

SURVEY OF HISTORICAL DATABASES

WITH LONGITUDINAL MICRO-DATA

The second questionnaire

For more information about this questionnaire or questions about entering specific information, please contact Kees Mandemakers (kma@iisg.nl) and/or Tatiana Moisseenko (tatiana.moisseenko@iisg.nl)

Table of Contents:

- | | | |
|-------|---|---------------------------------------------------|
| I. | A | General (identifying) information about databases |
| II. | A | Contact information |
| III. | A | Sources: core characteristics |
| IV. | A | Database: core characteristics |
| V. | A | Publications and reports |
| VI. | B | Observations |
| VII. | B | Sampling design and procedures |
| VIII. | B | Data collection |
| IX. | B | Linkage process |
| X. | C | Sources: the main characteristics of every source |

The questionnaire comprises three sections:

Section A includes the questions related to the most general and important information identifying the content, scope and provenance of the databases and the information about their creators.

Section B contains more specific and detailed questions about databases, such as the period(s) of observation, sampling design and procedures, data collection, linkage process and others.

Section C contains detailed questions about sources used for the databases: their type, scope, content, state of preservation, etc.

Section A

I. General (identifying) information about the database

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. Title of the database | POPLINK |
| 1.a. Subtitle , which brings meaning to the title (scope, place, time period): | |
| 2. Abbreviation | DDB (POPLINK) |
| 3. Links to website(s): | |
| 3.a. Homepage | http://www.ddb.umu.se/english/database/poplink/ |
| 3.b. Get to data | http://www.ddb.umu.se/english/service/order-data-retrieval |
| 4. Abstract: describes content of the database. Max. length: 300 words Please indicate: <ul style="list-style-type: none">◦ Scope and main goal◦ Time and territory covered by data◦ Sample strategy◦ Main sources | <i>Scope and main goal</i> The database POPLINK is one of the world's most information-dense historical population databases: it contains life-course data for 340 000 individuals and 1,2 million records between 1900 and 1950. The Demographic Data Base has digitized the Skellefteå and Umeå regions up until 1950 to enable linkage to modern registries and biobanks. By combining POPLINK with the Skellefteå regions in POPUM, the individual-level linkage between historical and modern data is made available and |

| | |
|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | <p>covers almost 400 years.</p> <p><i>Sample strategy</i> Complete registration of parish registers where individuals are followed during their presence within the included parishes. When choosing the target population, an important feature for maximizing the number of complete genealogies over generations is a low mobility. Homogeneity is vital in the medical sciences, as it will help minimize confounding by population stratification, reduce selection bias and facilitate correct imputations of genotypes. The coastal area of Västerbotten has until the early 20th century been characterized by a very low population turnover. There is an also optimal possibility of linking large modern registries covering the whole area.</p> <p><i>Time and territory covered by data</i> -Skellefteå region (seven parishes in northern Sweden), 1620 – 1950 -Umeå region (two parishes in northern Sweden), c. 1900-1950</p> <p><i>Main sources</i> Parish registers such as catechetical registers, birth and baptism registers, banns and marriage registers, migrations registers, and death registers.</p> |
|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

| | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|
| <p>5. Keywords: Please use the recommended keywords if they are applicable: <i>demography, life course, census, church register, civil certificates, population register, history, social science, genetics, migration, occupations.</i></p> | Demography, life course, parish registers, church records, genetics, occupations, migration, population history |
| Please add your own keywords, if you have data not covered by the recommended terms. | Epidemiology, literacy, fertility, mortality, family intergenerational |

| | |
|-------------------------------------------------------------------------|---------------|
| 6. Citation: Indicate how you want others to cite your database. | DDB (POPLINK) |
|-------------------------------------------------------------------------|---------------|

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------|--------------------------------|
| 7. IDS compatible: Indicate with <i>Yes</i> or <i>No</i> whether the database is IDS compatible, if <i>Yes</i> , please specify. | Yes, the time period pre 1900. |
|-----------------------------------------------------------------------------------------------------------------------------------------|--------------------------------|

| | |
|--------------------------------------------------------------------------------------|---------------------------------------------------------------------|
| 8. Has the database already been completed or it is still under construction? | Under construction. |
| 8.a. If completed, please indicate the years of its construction? | The completed parts was constructed between 2009-2015 |
| 8.b. If under construction, please indicate, when it is planned to complete it? | There is no clear date for completion. |
| 8.c. Please add a brief description of future plans for the database. | New parishes will be added as to increase the registered population |

II. Contact information

| | |
|---------------------------------------------|-------------------------------------------------------------------------------------------------------|
| 1. Name of institute or organisation | Demographic Data Base, Umeå University |
| 1.a. Website | http://www.ddb.umu.se/english/?languageId=1 |
| 1.b. Location: city, country | Umeå, Sweden |
| 1.c. Postal address | Umeå University, SE 901 87 Umeå, Sweden |

| | |
|-------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| 1.d. Phone | +46907865000 |
| 2. Name of primary responsible person | Anders Brändström |
| 2.a. His/her email address | Anders.brandstrom@ddb.umu.se |
| 2.b. Postal address | Umeå university, 901 87 Umeå, Sweden |
| 2.c. Phone | +46 90 7866063 |
| 3. Administrative information | |
| 3.a. When this form was filled? | February 17, 2015 |
| 3.b. Who did it? | Annika Westberg |
| 4. Main economic funding (Name of organization(s) who made the grants /sustain it) | Umeå university Swedish Research Council |

III. Sources: core characteristics

| 1. Type of the sources. | | | | | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------|---------------|-------------------|-----------------|------------------------------|
| Indicate how many sources were used for the database and what kind (register, census, certificates ...). Please enter <i>Yes</i> or <i>No</i> and the time period for the main sources. In case of other sources, not listed below, please add their type and specify their main characteristics. | | | | | |
| Detailed questions about the characteristics of all core sources are in section C. | | | | | |
| | <i>Type of source</i> | <i>Yes/No</i> | <i>Start year</i> | <i>End year</i> | <i>Explanations:</i> |
| 1. | Baptisms | Y | 1630 | 1950 | Include births and baptisms. |
| 2. | Marriages from church/parish registers | Y | 1700 | 1950 | |
| 3. | Burials | Y | 1620 | 1950 | Include deaths and burials. |
| 4. | Population registers | Y | 1720 | 1950 | |
| 5. | Civil birth certificates | N | | | |
| 6. | Civil marriage certificates | N | | | |
| 7. | Civil death certificates | N | | | |
| 8. | Population Census | N | | | |
| 9. | Nominative lists | N | | | |
| 10. | Military draft records | N | | | |
| 11. | Other: | | | | |

IV. The database: core characteristics

| 1. Period covered by data: give first and last year of date, if possible | | 1620 – 1950 (Skellefteå region) 1900 – 1950 (Umeå region) | | | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------|------------------------------------------------------------------------------------------------------------------------------|------------------------|-------------------------------------------------------------------------------------------------------------|--|
| 2. Territory covered by data | | Two geographically connected regions in northern Sweden: Skellefteå region (seven parishes) Umeå region (two parishes) | | | |
| 3. Geographical characteristic: local, regional, national, cross-national | | Regional | | | |
| 4. Units of observation. Please enter <i>Yes</i> or <i>No</i> for each unit, which forms the sample, the number of units and write explanations/comments. Add other units if they are not listed below, for them explanations are especially important. | | | | | |
| | <i>Units of observation:</i> | <i>Yes /No</i> | <i>Number of units</i> | <i>Explanations:</i> | |
| 1. | Individuals | Y | | | |
| 2. | Married couples | Y | | | |
| 3. | Families | Y | | | |
| 4. | Households | Y | | It might be difficult to identify households. | |
| 5. | Farms | Y | | Depends on how the population register was kept. It can differ from parish to parish and from time to time. | |

| | | | | |
|----|--------------|---|--|--|
| 6. | Institutions | N | | |
| 7. | Other | | | |

| | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 5. Variables per unit included in the database | |
| <p><u>On individuals:</u> <i>Data of birth and dead, age, gender, marital status, religion, occupation, migration, relationship, etc.</i> Please add more variables, if they are not in the list</p> | Dates on birth, baptism, death, burial and marriage. Legitimacy, age, gender, marital status, occupations with HISCO-coding, cause of death (ICD-10-coded), migration, relationships, literacy, smallpox vaccination. |
| <p><u>On households:</u> <i>Type of household, children present, age and number of children, etc.</i> Please add more variables, if they are not in the list</p> | Households might be difficult to identify. Families, including number of children (both biological and non-biological), are identified by relation and place of residence. |

| | |
|---------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 6. Kinship relations: | |
| 6.a. How is kinship recorded in the database? | A specific table contains information about related individuals. Given relations are to parents, partners and children. From this table sib ship groups can be created and families followed over generations. |
| 6.b. How deep (number of generations) is kinship information going? | Up to fifteen generations. |

| | |
|------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| 7. Completeness | |
| 7.a. Are all variables from the sources included in the database? | Yes |
| 7.b. Are all individuals who lived in the households of the sample recorded? | Yes, but it is difficult to connect servants and farm hands to a certain family. All individuals that have presence in the parish are recorded. |

| | |
|----------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>8. Current data representation: Database Software (e.g. MySql, MsSql, Access, please specify</p> | <p>INDIKO: Web tool for extracting and visualizing data (mainly visualizing) for the c 1650-1900: http://www.ddb.umu.se/english/service/indiko--parish-registers-on-the-web/?languageId=1</p> <p>DDB library: a set of standardized java methods for analysis and data extraction.</p> <p>CoreLink: computerized record linkage software.</p> <p>PERSONA: a new open source software for digitizing longitudinal data will be ready for use in late 2015 (http://www.ddb.umu.se/tjanster/v42---utveckling-i-forskningens-tjanst/)</p> |
|----------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

| | |
|--------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| 9. Access conditions: | |
| 9.a. How does a user get access to the database? | By contact with DDB. |
| 9.b. What are the conditions and restrictions? | Approval from an ethical committee is required for data from the period 1900-1950. Conditions and restrictions are also defined by contract. |

V. Publications and reports

1. Main publications about the database itself (max. 5)

- Edvinsson, Sören. "The Demographic Data Base at Umeå University - a resource for historical studies". In P. H. Hall, R McCaa and G. Thorvaldsen, *Handbook of International Historical Microdata for Population Research*, Minnesota Population Center 2000.
- Johansson, Egil. "Church Records - Part I: From Orality to Reading Tradition. Church Records - Part II: Baptism, Teaching to Observe, and the Demographic Data Base (DDB). Opening Reflections". In *Interchange*, vol. 34, no. 2 & 3, 2003.
- Nilsdotter Jeub, Ulla. *Parish Records. 19th Century Ecclesiastical Registers*. Demografiska databasen, Umeå 1993.
- Vikström, Pär, Edvinsson, Sören & Brändström, Anders. "Longitudinal databases – sources for analyzing the life course. Characteristics, difficulties and possibilities". *History and Computing* 2002 vol. 14
- Wisselgren, Maria, Edvinsson, Sören, Berggren, Mats & Larsson, Maria. "Testing Methods of Record Linkage on Swedish Censuses" in *Historical Methods* 2014, vol. 47, pp. 138-151.

2. Main or exemplary publications on research based on the database (max. 5)

- Egerbladh, Inez. & Bittles, Alan H. Socioeconomic, demographic and legal influences on consanguinity and kinship in northern coastal Sweden, 1780-1899. *Journal of Biosocial Science*, 22, pp. 1-23, 2011.
- Edvinsson, Sören, Brändström, Anders, Rogers, John & Broström, Göran. "High Risk Families: the unequal distribution of infant mortality in nineteenth century Sweden". *Population Studies*, vol. 59, 2005:3, 321-337.
- Engberg Elisabeth. "Boarded out by auction: poor children and their families in nineteenth-century northern Sweden", *Continuity and Change* 2004:19(3), pp. 431-457.
- Maas, Ineke & van Leeuwen, Marco H.D. "Industrialization and Intergenerational Mobility in Sweden". In *Acta Sociologica* 45: 179194, 2002.
- Vikström, Lotta. "Identifying dissonant and complementary data on women through the triangulation of historical sources", *International Journal of Social Research Methodology* 2010: vol.13, no. 3, pp. 211-221.

Section B

contains more specific and detailed questions about databases, such as the period(s) of observation, sampling design and procedures, data collection, linkage process and others.

VI. Observations

| | |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. How do individuals enter observation? | Birth, start of registration, in-migration. |
| 2. How do individuals leave observation? | Death, end of registration, out-migration |
| 3. How do households enter observation? | Only individuals are registered |
| 4. How do households leave observation? | Only individuals are registered |
| 5. Are some entry or exit dates unknown? | Only in rare cases. Mainly for older periods (i.e. 18th century). |
| 6. Are some entry or exit dates estimated? | Sometimes only year is given. |
| 7. Can observations be linked to geographic locations? | Yes |
| 8. Are the dates and locations of movements within the observation area recorded? | Yes |
| 9. Are all individuals who lived in selected households recorded? (Selection on basis of the sample or because sampled individuals are living in households) | Yes |
| 10. Are there related observations that are not included in the database? | Explicit information on related persons not present in the parish is included in the database (for example "daughter of farmer Nils Olsson", or "Farmers daughter"). |

VII. Sampling design and procedures: how was sample(s) defined?

| | |
|-----------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. Source(s): Which source forms the basis for the sample | Full count of parish registers for parishes selected by the research community. Parishes are grouped in two main regions. Individuals are followed during their presence within these regions |
| 2. Sampling units: Households, individuals, regions... | Complete registration |
| 3. Variables used for selection: Age, gender, marital status, other | Complete registration |
| 4. Selection method: Random, stratified random, total count, clustered, other | Full count |

VIII. Data collection

| | |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| 1. Data collection period: When the data was collected and transcribed? | The data from 1900-1950 was transcribed 2009-2015. |
| 2. Data collection method: Public digital register, transcription, other | Transcription from scanned original sources |
| 2.a. If transcription, how was the transcription done: <ul style="list-style-type: none"> ◦ By individuals ◦ From scanned sources ◦ From LDS's microfilms ◦ Automatic controls | By individuals from scanned original sources |
| 2.b. How was the checking of the transcription done? For example, by proof reading? | Automatic checks when transcribing, random sample checked by proof reading. |
| 2.c. When was it done? | At time of registration |
| 2.d. Purpose of the transcription: please indicate <ul style="list-style-type: none"> ◦ LDS ◦ Research ◦ Genealogy | Research |

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------|
| 3. Control methods by researcher: e.g. Internal consistencies such as a death cannot happen before a birth of the same person | Consistencies are checked by logical controls and computer programmes. |
| 4. Data collection staff: Please indicate the number of people and their position (member of the project, free-lancer, other) | |

IX. Linkage process

| | | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------|
| 1. Linkage: Which sources and units of observation have been linked: (e.g. birth/baptisms and death/burials,...)? | Births/Baptisms –Y Marriages -Y Deaths/Burials -Y Population registers - Y | |
| 2. Documentation of linking: | | |
| 2.a. Programme, manually, ... | We use a combination of computerized and manual linkage and link in three steps. First within the closest geographical unit (parish), then we link relations with parents and children and finally within a bigger geographical unit. | |
| 2.b. Name of software if used (and its parameters) | Software: CoreLink, Relink and RegLink for computerized linkage. ManLank and SirLink as computerized aid when linking was done manually | |
| 3. What are the rules for linking? Flags definition (list them: age, name, extra knowledge ...) | Several different rules are used during computerized linkage. Key variables are date of birth, sex, first name and last name. All links are logged to be traceable. | |
| 4. How each reconstructed person is traceable to the original sources /transcribed data? | Volume, page and row in the original sources are recorded for each individual | |
| 5. How is linkage represented in the database? For example, do all occurrences of an individual include a universal identification number (ID)? Or are records linked in another way? | Every individual has a unique identification number. Every record has a unique identification number and is linked to individuals through the unique person identification number. | |
| 6. Linkage percentage | 97-98% | |
| 7. Quality of linkage (own evaluation) | 100% | |
| 8. What reference/coding systems have been linked to the data? For example, occupational titles (like HISCO), locations (including geo-referenced systems). Please indicate the name of the system and how it was used. (Yes, No, Partly). | | |
| Y/N/P | Reference system | Explanations: |
| Y | Occupational titles: | Own coding system and HISCO |
| Y | Locations (including geo-referenced systems): | |
| Y | Religion, civil status etc.: | |
| Y | Other: | Cause of death according to ICD-10 |

Section C

contains detailed questions about sources used for the databases: their type, scope, content, state of preservation, etc.

Please answer the questions about all the sources used for the database, but do it in a separate form for every type of the source.

X. The main characteristics of the source (per every type of the source)

| | |
|-------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------|
| 1. Official name of the source and its English translation | Parish registers |
| 2. Purpose of the source: | |
| 2.a. Why was this source created? | Keeping track of vital events, taxation purposes, populations ability to read, control of migration |
| 2.b. Who created it? | Swedish Lutheran Church |
| 3. Scope: What group of the population was documented in this source? | Complete population |
| 4. Time period: When the information of the sources was recorded? Please indicate the start and the end date. | 1689-1989 |
| 5. Geographical area: What territory is covered by the source? | Sweden |
| 6. Content: What was recorded? | |
| 7. Language of written material: original sources and documentation | Swedish |
| 8. Preservation and storage: | |
| 8.a. Completely preserved | Yes |
| 8.b. Partially destroyed by personnel according to systematic criteria | |
| 8.c. Partially destroyed or damaged for other reasons | |
| 8.d. Reorganized by producer of the source | |
| 8.e. Reorganized by record linkage procedures | |
| 8.f. Where the original records are stored (name of the archive or institution)? | Swedish National Archives |
| 9. Documentation: | |
| 9.a. Completely documented and accessible by: | Swedish National Archives |
| 9.b. Partially documented and accessible by: | |
| 9.c. No documentation, but accessible by: | |