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:

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.

<u>**Section A**</u>

I.   General (identifying) information about the database

1. Title of the database	POPUM
1.a. Subtitle, which brings meaning to the title (scope, place, time period):	
2. Abbreviation	DDB (POPUM)

3. Links to website(s):	
3.a. Homepage	http://www.ddb.umu.se/english/database/the-database-popum
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:
◦ Scope and main goal
◦ Time and territory covered by data
◦ Sample strategy
◦ Main sources | *Scope and main goal*
The database POPUM is one of the world's most information-dense historical population databases. It contains information about 660 000 individuals and almost 5 million records, which cover the period of 1620-1900. In this database we have linked individual records from parish registers such as catechetical registers, birth and baptism registers, banns and marriage registers, migrations registers, and death registers.
Original goal is to digitize parish registers from |

	selected Swedish parishes and make them available for researchers. *Sample strategy* Complete registration of parish registers for parishes selected by the research society. Parishes grouped in four main regions. Individuals followed during presence within the included parishes. *Time and territory covered by data* Skellefteå region (seven parishes in northern Sweden), Sundsvall region (eighteen parishes in mid-northern Sweden), Linköping region (thirty-six parishes in southern Sweden) and Northern inland region (eleven parishes). 1620 – 1900 *Main sources* Parish registers such as catechetical registers, birth and baptism registers, banns and marriage registers, migrations registers, and death registers.

5. Keywords: Please use the recommended keywords if they are applicable: *demography, life course, census, church register, civil certificates, population register, history, social science, genetics, migration, occupations.*	Demography, life course, church register, history, social science, migration, occupations
Please add your own keywords, if you have data not covered by the recommended terms.	Intergenerational, mortality, fertility, family, literacy, epidemiology

6. **Citation:** Indicate how you want others to cite your database.	DDB (POPUM)

7. IDS compatible: Indicate with *Yes* or *No* whether the database is IDS compatible, if *Yes*, please specify.	Yes, a part of it is in IDS format (Skellefteå rural parish) and the remaining can be transformed into IDS.

8. Has the database already been completed or it is still under construction?	
8.a. If completed, please indicate the years of its construction?	The completed parts were constructed between 1973-2015
8.b. If under construction, please indicate, when it is planned to complete it?	Under construction. There is no date for when it will be completed.
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	

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, SE 90187 Umeå, Sweden
2.c. Phone	+46 90 7866063

3. Administrative information	

3.a. When this form was filled?	Feb 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 *Yes* or *No* 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.

	Type of source	*Yes/No*	*Start year*	*End year*	*Explanations:*
1.	Baptisms	Y	1630	1900	Include births and baptisms. Mainly late 18th and 19th century
2.	Marriages from church registers	Y	1700	1900	Mainly late 18th and 19th century.
3.	Burials	Y	1620	1900	Include deaths and burials. Mainly late 18th and 19th century.
4.	Population registers, maintained by church	Y	1720	1900	Mainly late 18th and 19th century.
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 - 1900
2. Territory covered by data	Sweden: Skellefteå region (seven parishes in northern Sweden), Sundsvall region (eighteen parishes in mid-northern Sweden), Linköping region (thirty-six parishes in southern Sweden) and Northern inland region (eleven parishes).
3. Geographical characteristic: local, regional, national, cross-national	Regional

4. Units of observation. Please enter *Yes* or *No* 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.

	Units of observation:	Yes /No	Number of units	*Explanations:*
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	
<u>On individuals:</u> *Data of birth and dead, age, gender, marital status, religion, occupation, migration, relationship,* etc. Please add more variables, if they are not in the list	Gender, age, dates of birth, baptism, death, burial and marriage, legitimacy, age, gender, marital status. Presence in the parish, participation in holy communion, literacy, delinquency, smallpox vaccination, migration, relationships (biological and non-biological), cause of death with ICD-10 coding, occupations with HISCO coding.
<u>On households:</u> *Type of household, children present, age and number of children, etc*. Please add more variables, if they are not in the list	Family composition, including number of children and their age are identified by relation and place of residence. It is difficult to identify with certainty servants, farm hands etc. in the family.

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 eleven 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.

8. Current data representation: Database Software (e.g. MySql, MsSql, Access, please specify	INDIKO: web tool for extracting and visualizing data (mainly visualizing).
	DDB library: a set of standardized java methods for analysis and data extraction.
	CoreLink: computerized record linkage software.
	PERSONA: a new open source software for digitizing longitudinal population data will be ready for use in late 2015 (http://www.ddb.umu.se/tjanster/v42---utveckling-i-forskningens-tjanst/)

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?	Conditions and restrictions are 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, number 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, p 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, migration.
2. How do individuals leave observation?	Death, end of registration, 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	Complete registration of parish registers for parishes selected by the research community. Parishes are grouped in four 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	Total count

VIII.   Data collection

1. Data collection period: When the data was collected and transcribed?	1973 – 2015
2. Data collection method: Public digital register, transcription, other	Transcription from scanned original sources
2.a. If transcription, how was the transcription done: ◦ 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 and random sample checked by proof reading
2.c. When was it done?	At time of registration
2.d. Purpose of the transcription: please indicate ◦ 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 control and computer programmes.
4. Data collection staff: Please indicate the number of people and their position (member of the project, free-lancer, other**)**	

<div align="center">

IX.   Linkage process

</div>

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, RelLink and RegLink for computerized linkage. ManLank and SirLink as computerized aid when linking 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 (ID). 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 <u>separate form</u> 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	Kyrkoböcker Church records
2. Purpose of the source:	
2.a. Why was this source created?	Keeping track of vital events, taxation purposes, population's ability to read, and 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?	The whole country
6. Content: What was recorded?	Birth and baptism, marriage, migration, death and burial, church
7. Language of written material: original sources and documentation	Swedish
8. Preservation and storage:	
8.a. Completely preserved	
8.b. Partially destroyed by personnel according to systematic criteria	
8.c. Partially destroyed or damaged for other reasons	X
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)?	The National Archives
9. Documentation:	
9.a. Completely documented and accessible by:	The National Archives
9.b. Partially documented and accessible by:	
9.c. No documentation, but accessible by:	