

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	MOSAIC PROJECT
1.a. Subtitle , which brings meaning to the title (scope, place, time period):	
2. Abbreviation	Mosaic
3. Links to website(s):	
3.a. Homepage	http://censusmosaic.org
3.b. Get to data	http://censusmosaic.org/web/data
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>Original goal</i> The Mosaic project identifies, gathers, harmonizes, and distributes surviving historical census microdata for Europe and beyond. It aims at building a comprehensive and detailed resource for the historic study of populations by historians, demographers, economists, and other researchers. The Mosaic project builds on historical integrated census microdata projects such as the Integrated Public Use Microdata Series (IPUMS) and the North Atlantic Population Project (NAPP) . Records are distributed in the same integrated format, so that historical comparisons can be made across time and space. <i>Sample definition</i>

	<p>A) Data provided by Mosaic partners: Sample definition as used in previous projects by various people – no consistent definition across all data files.</p> <p>B) Samples produced by Mosaic: Samples are based on surviving census material and try to cover the whole area of the census in question. Sample definitions differ and are provided with individual data files.</p> <p><i>Geographic area under observation</i> Europe, best coverage until now for Central and East Central Europe</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>	Census, demography, history
Please add your own keywords, if you have data not covered by the recommended terms.	

<p>6. Citation: Indicate how you want others to cite your database.</p>	Mosaic database
--	-----------------

<p>7. IDS compatible: Indicate with <i>Yes</i> or <i>No</i> whether the database is IDS compatible, if <i>Yes</i>, please specify.</p>	No
---	----

<p>8. Has the database already been completed or it is still under construction?</p>	It is still under construction
8.a. If completed, please indicate the years of its construction?	
8.b. If under construction, please indicate, when it is planned to complete it?	It is intended to enhance the database regularly.
8.c. Please add a brief description of future plans for the database.	Adding additional country samples in the years 2015-2016.

II. Contact information

<p>1. Name of institute or organisation</p>	Max Planck Institute for Demographic Research, Rostock / Department of Demography and Center for Demography and Economics of Aging (CEDA), University of California, Berkeley / Max Plank Institute for Social Anthropology, Halle-Saale / University of Graz
1.a. Website	www.censusmosaic.org
1.b. Location: city, country	Rostock, Germany Berkeley, USA Halle, Germany Graz, Austria
1.c. Postal address	Mozartgasse 3, 8010 Graz, Austria
1.d. Phone	

<p>2. Name of primary responsible person</p>	Siegfried Gruber Joshua R. Goldstein Mikołaj Szoltysek
2.a. His/her email address	si.gruber@uni-graz.at josh@demog.berkeley.edu Szoltysek@eth.mpg.de
2.b. Postal address	Mozartgasse 3, 8010 Graz, Austria

2.c. Phone	+43-316/380-2377
------------	------------------

3. Administrative information	
3.a. When this form was filled?	2015-02-20
3.b. Who did it?	Siegfried Gruber

4. Main economic funding (Name of organization(s) who made the grants /sustain it)	Max Planck Institute for Demographic Research
---	---

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.

	<i>Type of source</i>	<i>Yes/No</i>	<i>Start year</i>	<i>End year</i>	<i>Explanations:</i>
1.	Baptisms	N			
2.	Marriages from church registers	N			
3.	Burials	N			
4.	Population registers, maintained by church or state	N			
5.	Civil birth certificates	N			
6.	Civil marriage certificates	N			
7.	Civil death certificates	N			
8.	Population Census	Y	1430	1918	Mainly 18th and 19th centuries
9.	Nominative lists	N			
10.	Military draft records	N			
11.	Other: Status Animarum	Y	1430	1918	Mainly 18th and 19th centuries
12.	Other: Tax registers	Y	1430	1918	Mainly 18th and 19th centuries
13.	Other: Population enumeration listings	Y	1430	1918	Mainly 18th and 19th centuries

IV. The database: core characteristics

1. Period covered by data: give first and last year of date, if possible	1430 - 1918
---	-------------

2. Territory covered by data	Europe
-------------------------------------	--------

3. Geographical characteristic: local, regional, national, cross-national	local
--	-------

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	2,400,000	
2.	Married couples			Can be derived
3.	Families			Individuals are grouped by households
4.	Households			Individuals are grouped by households
5.	Farms			In rural areas households correspond generally to farms
6.	Institutions			
7.	Other			

5. Variables per unit included in the database	
On individuals: <i>Data of birth and dead, age, gender,</i>	age, first name, last name, marital status, occupation,

<i>marital status, religion, occupation, migration, relationship, etc.</i> Please add more variables, if they are not in the list	presence/absence, relationship to head, religion, gender, literacy, person weight
<i>On households: Type of household, children present, age and number of children, etc.</i> Please add more variables, if they are not in the list	size, group quarter status, household weight place, region, urban/rural
6. Kinship relations:	
6.a. How is kinship recorded in the database?	1. transcription of kinship title in census 2. coded in variable "relate"
6.b. How deep (number of generations) is kinship information going?	Households may contain two to four generations.
7. Completeness	
7.a. Are all variables from the sources included in the database?	No
7.b. Are all individuals who lived in the households of the sample recorded?	yes
8. Current data representation:	
Database Software (e.g. MySQL, MsSql, Access, please specify)	SPSS, distribution as CSV-files
9. Access conditions:	
9.a. How does a user get access to the database?	registration is necessary
9.b. What are the conditions and restrictions?	All registered users are granted a limited license to use and distribute this documentation and to use the accompanying data, subject to the following conditions: The data may not be redistributed. Publications and research reports based on the database must cite it appropriately.

V. Publications and reports

1. Main publications about the database itself (max. 5)
 - Mikołaj Szotłysek and Siegfried Gruber, Mosaic: recovering surviving census records and reconstructing the familial history of Europe. In: The History of the Family, forthcoming.
2. Main or exemplary publications on research based on the database (max. 5)
 - Siegfried Gruber and Mikołaj Szotłysek (2015) The patriarchy index: a comparative study of power relations across historical Europe. In: The History of the Family, [DOI: 10.1080/1081602X.2014.1001769](https://doi.org/10.1080/1081602X.2014.1001769)
 - Péter Öri, Levente Pakot (2014) Residence patterns in nineteenth century Hungary: Evidence from the Hungarian Mosaic Sample. Working Papers on Population, Family and Welfare, No. 20, Hungarian Demographic Research Institute, Budapest.
 - Siegfried Gruber and Mikołaj Szotłysek (2012) Stem Families, Joint Families, and the European Pattern: What kind of a Reconsideration Do We Need? In: Journal of Family History, 37:1, 105-125. [DOI: 10.1177/0363199011428124](https://doi.org/10.1177/0363199011428124)
 - Siegfried Gruber, Rembrandt D.Scholz, and Mikołaj Szotłysek (2011) Real and Synthetic Household Populations and Their Analysis: An Example of Early Historical Census Microdata (Rostock in 1819). In: Historical Methods: A Journal of Quantitative and Interdisciplinary History, 44:2, 107-113. [DOI: 10.1080/01615440.2010.517508](https://doi.org/10.1080/01615440.2010.517508)
 - Mikołaj Szotłysek, Siegfried Gruber, Barbara Zuber-Goldstein, Rembrandt Scholz (2010) Living arrangements and household formation in the crucible of social change: Rostock 1867-1900. Rostock, MPIDR Working Paper WP-2010-036.

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?	upon being included in a census
2. How do individuals leave observation?	no leaving
3. How do households enter observation?	upon being included in a census
4. How do households leave observation?	no leaving
5. Are some entry or exit dates unknown?	no entry or exit dates
6. Are some entry or exit dates estimated?	no entry or exit dates
7. Can observations be linked to geographic locations?	Yes
8. Are the dates and locations of movements within the observation area recorded?	No
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?	in some census lists additional information is provided as remarks (e.g. migration)

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

1. Source(s): Which source forms the basis for the sample	Sample definition A) Data provided by Mosaic partners: Sample definition as used in previous projects by various people – no consistent definition across all data files. B) Samples produced by Mosaic: Samples are based on surviving census material and try to cover the whole area of the census in question. Sample definitions differ and are provided with individual data files.
2. Sampling units: Households, individuals, regions...	settlements
3. Variables used for selection: Age, gender, marital status, other	
4. Selection method: Random, stratified random, total count, clustered, other	different selection methods depending on different samples

VIII. Data collection

1. Data collection period: When the data was collected and transcribed?	in the Mosaic project since 2011, data donated to the project has been collected and transcribed since the 1970s
2. Data collection method: Public digital register, transcription, other	scanning, transcription
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 sources, publications, Xerox copies, original census material in archives, partly automatic controls
2.b. How was the checking of the transcription done? For example, by proof reading?	sample proof reading plus logical cross checks
2.c. When was it done?	after transcription

2.d. Purpose of the transcription: please indicate <ul style="list-style-type: none"> ◦ LDS ◦ Research ◦ Genealogy 	research genealogy (for some donated data files)
3. Control methods by researcher: e.g. Internal consistencies such as a death cannot happen before a birth of the same person	completeness of codes, internal consistencies (e.g. a spouse must be married, children must be at least 14 years younger than parents)
4. Data collection staff: Please indicate the number of people and their position (member of the project, free-lancer, other)	member of the project (25) free-lancer (about 20) other (data donated to Mosaic, number unknown)

IX. Linkage process

1. Linkage: Which sources and units of observation have been linked: (e.g. birth/baptisms and death/burials...)?	no linkages yet	
2. Documentation of linking:		
2.a. Programme, manually, ...		
2.b. Name of software if used (and its parameters)		
3. What are the rules for linking? Flags definition (list them: age, name, extra knowledge ...)		
4. How each reconstructed person is traceable to the original sources /transcribed data?		
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?		
6. Linkage percentage		
7. Quality of linkage (own evaluation)		
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:
P	Occupational titles:	Occhisco used as in NAPP
Y	Locations (including geo-referenced systems):	latitude, longitude
Y	Religion, civil status etc.:	Religion: own system, similar to NAPP marital status: compatible to MARST in NAPP
Y	Other:	Relationship to head, sex, literacy, urban/rural, presence, and group quarter: based on codes of IPUMS and NAPP.

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.

Section C cannot be answered like this because the information is specific for each source and we have currently 249 entries in our project administration database.

Remark of Mag. Dr. Siegfried Gruber, Karl-Franzens-Universität Graz

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

1. Official name of the source and its English translation	
2. Purpose of the source:	
2.a. Why was this source created?	
2.b. Who created it?	
3. Scope: What group of the population was documented in this source?	
4. Time period: When the information of the sources was recorded? Please indicate the start and the end date.	
5. Geographical area: What territory is covered by the source?	
6. Content: What was recorded?	
7. Language of written material: original sources and documentation	
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	
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)?	
9. Documentation:	
9.a. Completely documented and accessible by:	
9.b. Partially documented and accessible by:	
9.c. No documentation, but accessible by:	