

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	Aranjuez Database: Individual and family trajectories.
1.a. Subtitle , which brings meaning to the title (scope, place, time period):	1871-1970
2. Abbreviation	ADB

3. Links to website(s):	
3.a. Homepage	http://www.geps.es/
3.b. Get to data	N/A

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 covers much of the demographic transition in Aranjuez. All of the papers generated from the database thus far have dealt with one aspect or another of this process. <i>Time and territory covered by data</i> The data refer to the town of Aranjuez, located some 50 km south of Madrid. Over the period, the population of the town ranges from some 7500 in the 1870s, to around 12,000 in the early twentieth century, and upwards of 30,000 towards the end of the period. <i>Sample strategy</i>
---	---

	There is no sampling at all in the database. The research group was given access to the civil registration material for the entire period. The same holds true for the local population listings and the military heights. Since the material focuses on a small town, there is considerable loss of information regarding people who out-migrate. Assessing the effects of mobility on the individuals in the database is extremely difficult.
--	---

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>	Life course, civil register, social science, occupations
Please add your own keywords, if you have data not covered by the recommended terms.	Family history, individual trajectories

6. Citation: Indicate how you want others to cite your database.	Aranjuez Database: Individual and family trajectories.
---	--

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

8. Has the database already been completed or it is still under construction?	
8.a. If completed, please indicate the years of its construction?	Completed
8.b. If under construction, please indicate, when it is planned to complete it?	
8.c. Please add a brief description of future plans for the database.	

II. Contact information

1. Name of institute or organisation	Population and Society Research Group (GEPs)
1.a. Website	http://www.geps.es/
1.b. Location: city, country	Madrid, Spain
1.c. Postal address	Dep. Sociología II, Facultad de CC.PP. y Sociología, Universidad Complutense de Madrid, 28223 Madrid, Spain
1.d. Phone	

2. Name of primary responsible person	Alberto Sanz-Gimeno
2.a. His/her email address	asanzg@cps.ucm.es
2.b. Postal address	
2.c. Phone	

3. Administrative information	
3.a. When this form was filled?	06.02.2015
3.b. Who did it?	Alberto Sanz-Gimeno

4. Main economic funding (Name of organization(s) who made the grants /sustain it)	Comunidad de Madrid
---	---------------------

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				
2.	Marriages from church registers				
3.	Burials				
4.	Population registers, maintained by church or state				
5.	Civil birth certificates	Y	1871	1970	
6.	Civil marriage certificates	Y	1871	1970	
7.	Civil death certificates	Y	1871	1970	
8.	Population Census				
9.	Nominative lists	Y	1877	1975	Six different 'padrones': 1877, 1905, 1912, 1945, 1960, 1975
10	Military draft records	Y	1893	1970	Heights of local draftees
11.	Other:				

IV. The database: core characteristics

1. Period covered by data: give first and last year of date, if possible		1871-1970			
2. Territory covered by data		Aranjuez, Spain			
3. Geographical characteristic: local, regional, national, cross-national		local			
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.					
	Units of observation:	Yes /No	Number of units	Explanations:	
1.	Individuals	YES	68462		
2.	Married couples	YES	10893		
3.	Families	YES	40143		
4.	Households			Households can be followed over different local population listings, but only if the data are linked. This has not been done so far.	
5.	Farms				
6.	Institutions				
7.	Other				

5. Variables per unit included in the database

On individuals: *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

On households: *Type of household, children present, age and number of children, etc.*

Please add more variables, if they are not in the list

Data of birth and dead, age, gender, marital status, occupation, place of birth, cause of dead, relationship, family data.

6. Kinship relations:

6.a. How is kinship recorded in the database?	Immediate kin relations can be derived (often indirectly) from the local nominative listings of households and inhabitants. For those who are born in Aranjuez, kin ties can be imputed thanks to family reconstitution.
6.b. How deep (number of generations) is kinship information going?	Two and possibly three

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

8. Current data representation: Database Software (e.g. MySQL, MsSql, Access, please specify)	All data base construction and management has been done basically with MS Access.
---	---

9. Access conditions:	
9.a. How does a user get access to the database?	The request for the permission to use the database should be directed to Alberto Sanz-Gimeno (asanzg.ucm@gmail.com) or to David Reher. The additional mail for this purpose: info@geps.es
9.b. What are the conditions and restrictions?	Access to database is under permission and requires knowledge of Microsoft Access because data structure is not adapted.

V. Publications and reports

1. Main publications about the database itself (max. 5)
 - REHER, David, GONZÁLEZ QUIÑONES, Fernando R. y SANZ GIMENO, Alberto (2004), «Procesos de modernización y trayectorias de vida. Propuesta para el análisis sociodemográfico a partir de datos locales en España», *Reconstituição de famílias, fogos e estratégias sociais, Actas del VI Congreso de la Asociación de Demografía Histórica, 18-20 abril 2001, Castelo Branco*.
2. Main or exemplary publications on research based on the database (max. 5)
 - REHER, David-Sven y GONZÁLEZ QUIÑONES, Fernando R. (2003), «Do parents really matter? Child health and development in Spain during the demographic transition», *Population Studies*, 57, 1, págs. 63-75.
 - COLANTONIO, Sonia E., FUSTER, Vicente, SANZ GIMENO, Alberto y REHER, David S. (2008), «Factors related to inbreeding components from isonymy in an urban population: Aranjuez (Spain)», *Journal of Biosocial Science* 40, 2, págs. 239-246.
 - REHER, David-Sven y SANZ GIMENO, Alberto (2007), «Rethinking historical reproductive change: Insights from longitudinal data for a Spanish town», *Population and Development Review*, 33, 4, págs. 703-727.
 - REHER, David-Sven, ORTEGA, José Antonio y SANZ GIMENO, Alberto (2008), «Intergenerational transmission of reproductive traits in Spain during the demographic transition», *Human Nature*, 19, 1, págs. 23-43.
 - VAN POPPEL, Frans, REHER, David S., SANZ GIMENO, Alberto, SÁNCHEZ DOMÍNGUEZ, María y BEEKINK, Erik (2012), «Mortality decline and reproductive change during the Dutch demographic transition: Revisiting a traditional debate with new data», *Demographic Research*, 27, 11, págs. 299-338.

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?	Individuals enter observation when they appear in the data base. For those who are born in Aranjuez, this is at birth. For in-migrants this can happen at any time on any type of document.
2. How do individuals leave observation?	Individuals leave observation at death or when they disappear from the data base. We can look at this issue indirectly by establishing exactly until what date people are really under observation.
3. How do households enter observation?	
4. How do households leave observation?	
5. Are some entry or exit dates unknown?	Of course.
6. Are some entry or exit dates estimated?	Yes, see above.
7. Can observations be linked to geographic locations?	Yes
8. Are the dates and locations of movements within the observation area recorded?	Partly
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?	As said earlier, everyone is included in the sample. The appearance of an individual in the data base might only be once, but full information on him is registered. In other words, some individuals appear many times in the data base (on birth, marriage and death certificates, on population listings, on military conscription lists) while others may only appear once.

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

1. Source(s): Which source forms the basis for the sample	There is no sampling at all in the database.
2. Sampling units: Households, individuals, regions...	
3. Variables used for selection: Age, gender, marital status, other	
4. Selection method: Random, stratified random, total count, clustered, other	

VIII. Data collection

1. Data collection period: When the data was collected and transcribed?	1998-2002
2. Data collection method: Public digital register, transcription, other	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
2.b. How was the checking of the transcription done? For example, by proof reading?	Automatic and semi-automatic with other manual controls.

2.c. When was it done?	2001-2003
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	Internal consistencies between different events
4. Data collection staff: Please indicate the number of people and their position (member of the project, free-lancer, other)	12-14 persons (researchers and students)

IX. Linkage process

1. Linkage: Which sources and units of observation have been linked: (e.g. birth/baptisms and death/burials...)?	Births, marriages, deaths, military conscriptions and local censuses
--	--

Births / Baptisms	Y	
Marriages	Y	
Deaths/Burials	Y	
Population registers	N	
Census	N	
Nominative lists	Y	
Other:	Y	Heights of military conscripts

2. Documentation of linking:	
2.a. Programme, manually, ...	Programme and manually
2.b. Name of software if used (and its parameters)	Microsoft Access
3. What are the rules for linking? Flags definition (list them: age, name, extra knowledge ...)	Names, age, marital status, dates
4. How each reconstructed person is traceable to the original sources /transcribed data?	By means of an identification number (ID)
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?	All individuals have a personal identification number (ID). This has been the basis of what we call a <i>Relational Table</i> that contains the IDs of every person in the database who appears on any of the records pertaining to a given individual as well as their position in those records (for example, his parents on his birth, marriage or death certificate). In this way, all links between different IDs can be ascertained easily. The main data base, of course, contains several other smaller data bases that have been generated for specific analytical purposes. The origin of all of them is the <i>Relational Table</i> .
6. Linkage percentage	90%
7. Quality of linkage (own evaluation)	95%
8. What reference/coding systems have been linked to the data? For example, occupational titles (like HISCO),	Occupational titles (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:	Occupational titles: We have generated our own system. Very few classifications are used, mainly due to the fact that a large part of the town's population is often classified as 'day laborers'. It also includes the classification by HISCO.
Y	Locations (including geo-referenced systems):	Locations
Y	Religion, civil status etc.:	Civil status, literacy, sex...
Y	Other:	Cause of death

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	
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?	City of Aranjuez (Madrid)
6. Content: What was recorded?	
7. Language of written material: original sources and documentation	Spanish
8. Preservation and storage:	
8.a. Completely preserved	X
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)?	Juzgado Municipal – Archivo Historico de Aranjuez
9. Documentation:	
9.a. Completely documented and accessible by:	
9.b. Partially documented and accessible by:	
9.c. No documentation, but accessible by:	X