

SURVEY OF HISTORICAL DATABASES
WITH LONGITUDINAL MICRO-DATA

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	Portuguese Genealogical Repository
1.a. Subtitle , which brings meaning to the title (scope, place, time period):	Portuguese Genealogies (16th-20th)
2. Abbreviation	PORGENER
3. Links to website(s):	
3.a. Homepage	http://www.genealog.uminho.pt/ http://www.ghp.ics.uminho.pt/
3.b. Get to data	http://www.ghp.ics.uminho.pt/genealogias.html
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 	Several databases built from baptism, marriage and death records. These databases, in a process of integration, currently include over 1 million individuals, covering the North of Portugal and two islands of the Azores Archipelago.
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>	Demography; life course; church registers; population register; history; social science; genetics; migration; occupations.
Please add your own keywords, if you have data not	

covered by the recommended terms.	
6. Citation: Indicate how you want others to cite your database.	PORGENER, Universidade do Minho
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?	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?	Considering the extent of the project it is not possible to plan a date.
8.c. Please add a brief description of future plans for the database.	Partnerships with local municipalities to enlarge the database.

II. Contact information

1. Name of institute or organisation	PORGENER, Universidade do Minho
1.a. Website	http://www.genealog.uminho.pt/
1.b. Location: city, country	Guimarães, Portugal
1.c. Postal address	Campus de Azurém, 4800-058 Guimarães
1.d. Phone	253 510 579
2. Name of primary responsible person	Maria Norberta Amorim
2.a. His/her email address	mnsbamorim@gmail.com
2.b. Postal address	Campus de Azurém, 4800-058 Guimarães
2.c. Phone	253 510 579
3. Administrative information	
3.a. When this form was filled?	23/06/2015
3.b. Who did it?	Antero Ferreira
4. Main economic funding (Name of organization(s) who made the grants /sustain it)	Universidade do Minho

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	Yes	~1550	1910	
2.	Marriages from church registers	Yes	~1550	1910	
3.	Burials	Yes	~1550	1910	
4.	Population registers, maintained by church or state	Yes			
5.	Civil birth certificates				
6.	Civil marriage certificates				
7.	Civil death certificates				

8.	Population Census				
9.	Nominative lists	Yes			
10.	Military draft records	Yes			
11.	Other:				

IV. The database: core characteristics

1. Period covered by data: give first and last year of date, if possible		~1550 – 1910			
2. Territory covered by data		Portugal			
3. Geographical characteristic: local, regional, national, cross-national		Cross-national			
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	1.000.000	As the DB is genealogically linked, all the family links (and marriages) in the DB are relations between individuals.	
2.	Married couples	Yes			
3.	Families	Yes			
4.	Households				
5.	Farms				
6.	Institutions				
7.	Other				

5. Variables per unit included in the database		
<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		Burial procedures; Testaments...
<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		

6. Kinship relations:		
6.a. How is kinship recorded in the database?	Relations between individual records (family of origin - ancestors; new family – wife and descendants)	
6.b. How deep (number of generations) is kinship information going?	To the last extension of the collected information.	

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?	No. There are often gaps with the register of minors, under seven years.	

8. Current data representation: Database Software (e.g. MySql, MsSql, Access, please specify)	MsSql and Access.
---	-------------------

9. Access conditions:		
9.a. How does a user get access to the database?	Send an email to PORGENER.	
9.b. What are the conditions and restrictions?	Level of elaboration of the database; Authorisation of the author;	

	Justification of the interest in the DB;
--	--

V. Publications and reports

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

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?	By the first event registered, mainly by birth.
2. How do individuals leave observation?	By the last event registered, mainly by death or migration.
3. How do households enter observation?	By the first event registered, mainly by marriage.
4. How do households leave observation?	By the last event registered, mainly by death or migration.
5. Are some entry or exit dates unknown?	Yes.
6. Are some entry or exit dates estimated?	Yes.
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)	All individuals registered in the sources are recorded. There are problems with death registers of minors, under seven.
10. Are there related observations that are not included in the database?	Yes. Vicar, godfathers, witnesses are not included systematically.

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

1. Source(s): Which source forms the basis for the sample	
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?	We have begun the collection of data by 1970 and are still collecting data.
2. Data collection method: Public digital register, transcription, other	Transcription
2.a. If transcription, how was the transcription done:	Individuals; scanned sources; LDS's microfilms.

<ul style="list-style-type: none"> ◦ By individuals ◦ From scanned sources ◦ From LDS's microfilms ◦ Automatic controls 	
2.b. How was the checking of the transcription done? For example, by proof reading?	Proof reading; evaluation of the transcriber.
2.c. When was it done?	In the linkage process and in the production of scientific studies..
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	Yes. Controlled by the software.
4. Data collection staff: Please indicate the number of people and their position (member of the project, free-lancer, other)	37 .

IX. Linkage process

1. Linkage: Which sources and units of observation have been linked: (e.g. birth/baptisms and death/burials...)?	birth/baptisms and death/burials	
2. Documentation of linking:		
2.a. Programme, manually, ...	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?	Every information is referred to a source and a date.	
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?	There is a universal identification number.	
6. Linkage percentage	80% (rural 95% urban 75%)	
7. Quality of linkage (own evaluation)	Very good	
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:
	Occupational titles:	
Y	Locations (including geo-referenced systems):	Administrative codes; geo-referencing under development.
Y	Religion, civil status etc.:	Civil status
	Other:	

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	Church registers
2. Purpose of the source:	Baptism, Marriage and Death registers
2.a. Why was this source created?	
2.b. Who created it?	Catholic church
3. Scope: What group of the population was documented in this source?	Catholic (99% population)
4. Time period: When the information of the sources was recorded? Please indicate the start and the end date.	16 th – 20 th
5. Geographical area: What territory is covered by the source?	Portugal (mainly north and Azores Islands)
6. Content: What was recorded?	
7. Language of written material: original sources and documentation	Portuguese
8. Preservation and storage:	
8.a. Completely preserved	Yes (with minor gaps)
8.b. Partially destroyed by personnel according to systematic criteria	No.
8.c. Partially destroyed or damaged for other reasons	Yes.
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)?	Regional Public Archives
9. Documentation:	
9.a. Completely documented and accessible by:	
9.b. Partially documented and accessible by:	
9.c. No documentation, but accessible by:	