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:

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	HISTORICAL POPULATION DATABASE OF TRANSYLVANIA (1850-1914)
1.a. Subtitle, which brings meaning to the title (scope, place, time period):	-
2. Abbreviation	HPDT

3. Links to website(s):	
3.a. Homepage	**http://hpdt.granturi.ubbcluj.ro/**
3.b. Get to data	**http://82.208.133.141:6080/**

| 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 main aim of the *Historical Population Database of Transylvania* (HPDT) project is to build the first historical database of Transylvania's population, covering the period 1850-1914 (from the first modern census to the limit of vital registration research allowed by Romanian Law).
Sample strategy
Considering the chronological and financial limitations, as well as the state of the sources, the project aims to cover 5-10% of the population of Transylvania during the period studied (ca. 5.4 million births) thus providing 350-500.000 individual entries. |

	The envisaged project duration is 36 months, but the working plan is structured so as to allow the subsequent continuous development of the database. Twelve micro-zones have been established as starting areas for data gathering.
	The micro-zones are scattered across Transylvania, covering various types of geographical areas (mountains, hills, valleys), historical regions (former border regiments, former Szekely sees and Romanian districts), and commercial and economic centers (market towns, railroad knots). The selected localities usually host two or more parishes of different denominations and ethnicity. The bulk of sources is formed by parish registers of various denominations. *Main sources*
	Alternatively, the relevance of other sources will be tested, such as: military muster rolls, civil registration, obituaries, matrimonial courts, etc. The lack of censuses is due to the original documents not being preserved.

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.	-

6. Citation: Indicate how you want others to cite your database.	Historical Population Database of Transylvania

7. IDS compatible: Indicate with *Yes* or *No* whether the database is IDS compatible, if *Yes*, please specify.	YES. The sources component of the HPDT database was built having in mind the IDS version at the time being (August 2014), and we strongly intend to keep the other components, especially the central database, IDS-compatible.

8. Has the database already been completed or it is still under construction?	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?	2017
8.c. Please add a brief description of future plans for the database.	The current funding plan stops in April 2017. If no other funding plan will be available at that date, the core-team members will continue to work and develop the database, at a slower pace, as part of their institutional work assignments. If further funding will be attracted we intend to expand the database geographically and chronologically-wise, in order to cover other micro-areas from Transylvania starting from the end of the 18th century.

II. Contact information

1. Name of institute or organisation	Babeș-Bolyai University, Centre for Population Studies
1.a. Website	http://centre.ubbcluj.ro/csp
1.b. Location: city, country	Cluj-Napoca, Cluj
1.c. Postal address	68 Avram Iancu street, 400083
1.d. Phone	+40-264-599613

2. Name of primary responsible person	Prof. Ioan Bolovan
2.a. His/her email address	ioanbolovan62@gmail.com
2.b. Postal address	68 Avram Iancu street, 400083, Cluj-Napoca, Cluj, Romania
2.c. Phone	+40-264-599613

3. Administrative information	
3.a. When this form was filled?	17.03.2015
3.b. Who did it?	Luminița Dumănescu

4. Main economic funding (Name of organization(s) who made the grants /sustain it)	EEA grants & Norway Grants

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	yes	1850	1914	
2.	Marriages from church registers	yes	1850	1914	
3.	Burials	yes	1850	1914	
4.	Population registers, maintained by church or state	yes	1850	1914	At this time, censuses and populations registers maintained by church only.
5.	Civil birth certificates	-	-	-	
6.	Civil marriage certificates	-	-	-	
7.	Civil death certificates	-	-	-	
8.	Population Census	-	-	-	
9.	Nominative lists	-	-	-	
10	Military draft records	-	-	-	
11.	Other:	-	-	-	

IV. The database: core characteristics

1. Period covered by data: give first and last year of date, if possible	1850-1914
2. Territory covered by data	Pre-defined micro-zones from all parts of Transylvania
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	**Yes**	**Number of**	*Explanations:*

	observation:	/No	units	
1.	Individuals	Yes	70.000	Since the database is in its early phases, and we haven`t yet started the standardization procedures, it is difficult to estimate the actual number of individuals. At the end of March 2015 there were ca. 70.000 persons recorded by HPDT
2.	Married couples			-
3.	Families			-
4.	Households			-
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	*Date and place of birth, baptism, betrothal, marriage, death and burial, onomastic (including nickname), legitimacy at birth, age, gender, marital status, religion, occupation, residence, nobility, doctoral degree, relations to other persons in the database*
On households: *Type of household, children present, age and number of children, etc*. Please add more variables, if they are not in the list	

6. Kinship relations:	
6.a. How is kinship recorded in the database?	Kinship is recorded, mostly, as given by the sources (parish registers). We intend to build a genealogical component of the database, for general public use, but this will only be done by early 2017.
6.b. How deep (number of generations) is kinship information going?	At this point, usually one generation (parents and children). However, the particularities of the Romanian onomastic, which records filiations over 2-3 generations (X son of Y son of Z) make deeper kinship information available.

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	MySql

9. Access conditions:	
9.a. How does a user get access to the database?	Given the current state of development of the database, no access to third-party users is being allowed.
9.b. What are the conditions and restrictions?	After HPDT reaches a proper level of development we intend to allow the access of researchers, based on their specific interests and on the relevance of their research for the database promotion and development. The conditions and restrictions of access will be detailed at that time.

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 birth recorded in a parish register from a selected micro-zone
2. How do individuals leave observation?	By death recorded in a parish register from a selected micro-zone
3. How do households enter observation?	-
4. How do households leave observation?	-
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, if mentioned
9. Are all individuals who lived in selected households recorded? (Selection on basis of the sample or because sampled individuals are living in households)	-
10. Are there related observations that are not included in the database?	NO

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

1. Source(s): Which source forms the basis for the sample	Parish-registers (baptisms, betrothals, marriages, deaths, change in denomination, confirmations, blessings, vaccination)
2. Sampling units: Households, individuals, regions…	The sampling is being made on geographical criteria: the basic sampling unit is the parish.
3. Variables used for selection: Age, gender, marital status, other	The main selection criterion is represented by the state of the sources: parishes with the best-preserved sets of registers go first, followed by the neighbouring parishes, in the same micro-zone, whit less preserved sources.
4. Selection method: Random, stratified random, total count, clustered, other	Twelve micro-zones have been established as starting areas for data gathering according to our own methodology. The micro-zones are scattered across Transylvania, covering various types of geographical areas (mountains, hills, valleys), historical regions (former border regiments, former Szekely sees and Romanian districts), and commercial and economic centers (market towns, railroad knots). The selected localities usually host two or more parishes of different denominations and ethnicity. The bulk of sources is formed by parish registers of various denominations. The selected localities are going to be entirely included in the database.

VIII.   Data collection

1. Data collection period: When the data was collected and transcribed?	From November 2014 onward
2. Data collection method: Public digital register, transcription, other	Transcription
2.a. If transcription, how was the transcription done: ◦ By individuals ◦ From scanned sources ◦ From LDS's microfilms	By individuals

◦ Automatic controls	
2.b. How was the checking of the transcription done? For example, by proof reading?	Two levels of proof-reading: 1. cross-proof-reading by transcribers 2. proof-reading by experienced historians and palaeographers
2.c. When was it done?	November 2014 onward
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	Internal validation based on unicity, presence and event-date. In this stage of development the internal validation is mainly oriented towards
4. Data collection staff: Please indicate the number of people and their position (member of the project, free-lancer, other**)**	30 13 team-core members 8 data entry operators 9 volunteers (March 2015)

IX.   Linkage process

1. Linkage: Which sources and units of observation have been linked: (e.g. birth/baptisms and death/burials…)?	No sources or units of observation have been linked yet. The linkage process should start only in late 2015.
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:*
Y	Occupational titles:	HISCO
Y	Locations (including geo-referenced systems):	LAU-2 codes adapted
Y	Religion, civil status etc.:	Mostly NAPP and MOSAIC codes, with some adaptations
	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	Protocoale parohiale (RO) – Egyházi anyakönyvek (HU) – Parish Registers
2. Purpose of the source:	
2.a. Why was this source created?	Church registers of baptisms, betrothals, marriages, deaths, etc. have been created in order for the Church, and later on the State, to have a clear statistical image of the population movement in specific countries or regions.
2.b. Who created it?	The Church. In Transylvania, these registers have been created by the following denominations: Orthodoxies, Armenian Orthodoxies, Greek-Catholics, Catholics, Armenian Catholics, Calvinists, Lutherans, Unitarians, and Jews.
3. Scope: What group of the population was documented in this source?	The entire population
4. Time period: When the information of the sources was recorded? Please indicate the start and the end date.	The oldest preserved registers date back to 1614. The bulk of the parish registers was created starting with late 1700s. From 1895 on civil registration was instated, but the recordings in the parish registers continue up to this day. HPDT uses mainly information from the time period 1850-1914.
5. Geographical area: What territory is covered by the source?	The territory of the former Great Principality of Transylvania before 1867, covered by today counties of Brașov, Sibiu, Hunedoara, Alba, Covasna, Harghita, Mureș, Bistrița-Năsăud, Cluj, Sălaj.
6. Content: What was recorded?	The registers recorded baptisms, betrothals, marriages, deaths, changes in faith, blessings, and vaccination. The information about individuals varies greatly across the sources, from basic onomastics (name, surname, and nickname) to age, sex, place of birth and residence, legitimacy at birth, denomination and occupation.
7. Language of written material: original sources and documentation	The languages of the written sources are: Romanian, Hungarian, German, Latin, and Hebrew. The documentation is developed in Romanian and English.
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	YES. The main reason for which these sources

	have been destroyed over time is related to the negligence of the Church personnel and archive personnel
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 original records are stored at the County Archives Services in Transylvania (Brașov, Sibiu, Hunedoara, Alba, Covasna, Harghita, Mureș, Bistrița-Năsăud, Cluj, and Sălaj). Some records are being preserved, up to this day, at the parishes.
9. Documentation:	
9.a. Completely documented and accessible by:	
9.b. Partially documented and accessible by:	
9.c. No documentation, but accessible by:	No documentation, but accessible at the County Archives from Transylvania.