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.

Section A

I.   General (identifying) information about the database

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

3. Links to website(s):	
3.a. Homepage	http://www.karjalatk.fi/index.php?lang=en
3.b. Get to data	https://katiha.mamk.fi

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	The Karelian Database consists of data from parish registers in the Karelian region. *Original goal* The objectives of the Karelian Database Foundation are to promote access of demographic data for research purposes, to promote research on the people of the ceded Karelia region and to maintain and develop Karelian data archives. In all, at the moment the material on source database files available on Internet relates to about 5 million entries, each presenting data from one person or event. *Sample definition*

	The database includes whole material from e.g. registers of births in the period 1682-1950: a total number of 1260,000 children.
	Geographic area under observation
	Before the Second World War in Karelian region there were all together 51 Lutheran and 22 Orthodox Parishes.
	Main sources
	Data are collected from catechetical lists, lists of children, and registers of vital statistics (registers of births, marriages, migrations and deaths) from the period 1681-1949.
	Realized parts
	Data in the Katiha is from the period 1681-1950 and relates to about 10 million entries, but only a part (data older than 100 years and all death registers) of them is published on Internet.

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, church 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.	

7. IDS compatible: Indicate with *Yes* or *No* whether the database is IDS compatible, if *Yes*, please specify.	No

8. Has the database already been completed or it is still under construction?	The database is still under construction, but available to use. For instance the statistics form the database: https://katiha.mamk.fi/statistics.jsp Collecting the data is still going on
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 planned to complete the Karelian database until year 2020.
8.c. Please add a brief description of future plans for the database.	The recording work is still going on (already over 25 years). There are already about 10 million entries in the database, but it has been calculated that there are 1,5 million entries left.

.

II.   Contact information

1. Name of institute or organisation	Karelian Database Foundation
1.a. Website	http://www.karjalatk.fi/index.php?lang=en https://katiha.mamk.fi
1.b. Location: city, country	Mikkeli, Finland
1.c. Postal address	Savilahdenkatu 10 A, 50100 Mikkeli, Finland
1.d. Phone	

2. Name of primary responsible person	Satu Soivanen
2.a. His/her email address	kts@karjalatk.fi
2.b. Postal address	

2.c. Phone	

3. Administrative information	
3.a. When this form was filled?	on 23th February 2015
3.b. Who did it?	Satu Soivanen

4. Main economic funding (Name of organization(s) who made the grants /sustain it)	Mikkeli University of Applied Sciences maintains the database. http://www.mamk.fi/r_d/digital_archiving_and_eservices

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	1682	1950	about 70 parishes
2.	Marriages from church registers	Yes	1681	1950	about 70 parishes
3.	Burials	Yes	1688	1950	about 70 parishes
4.	Population registers, maintained by church or state	Yes	1737	1950	about 70 parishes
5.	Civil birth certificates	Yes	1920	1950	about 40 registers
6.	Civil marriage certificates	Yes	1920	1950	about 40 registers
7.	Civil death certificates	Yes	1920	1950	about 40 registers
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	1681 - 1950
2. Territory covered by data	The Old Eastern Finland, Karelia
3. Geographical characteristic: local, regional, national, cross-national	Regional, https://katiha.mamk.fi/map.jsp

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	1 200,000	Children in Birth registers
2.	Married couples	Y	620,000	Persons in Marriage registers
3.	Families		5 300,000	Persons in Catechetical lists (same members of the family could be found several times from Catechetical lists, depend on the period).
4.	Households			Depends on how the population register was kept. Can differ from parish to parish and from time to time.
5.	Farms	N		
6.	Institutions	N		
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	Data of birth and dead, age, gender, marital status, religion, occupation, migration, name, parish
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?	The kinship or relationship to the head of the household is mentioned in the Catechetical lists. From this, the relationship (parents, partners and children) to the research person can be derived.
6.b. How deep (number of generations) is kinship information going?	As far as previous or later generations are found in the Catechetical lists (1737-1950).

7. Completeness	
7.a. Are all variables from the sources included in the database?	Yes, except dates of participation in the communion
7.b. Are all individuals who lived in the households of the sample recorded?	Yes.
7.c. Are the main variables comparable?	They are quite comparable, but the generation born before 1850 is less comparable because the database is incomplete. For instance recording of the oldest Catechetical lists is still going on.

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

9. Access conditions:	
9.a. How does a user get access to the database?	Via web service
9.b. What are the conditions and restrictions?	The private data can only be seen with user account

V.   Publications and reports

1.   Main publications about the database itself (max. 5)
- Johanna Räisä and Mirja Loponen 2014: The modernization, migration and archiving of a research register http://purl.pt/26107/1/DLM2014_PDF/13%20-%20The%20modernization,%20migration%20and%20archiving%20of%20a%20research%20register.pdf
- Satu Soivanen 2011: Katiha Statistics Program: Statistical Requirements for the Data of a Demographic Database https://www.theseus.fi/bitstream/handle/10024/36955/Soivanen_Satu.pdf?sequence=1

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

- Päivi Happonen 2009: Two realities? The parish registers and the population registers in describing the Sortavala city population profile from the beginning of the 19th century until 1940. http://epublications.uef.fi/pub/urn_isbn_978-952-219-316-2/urn_isbn_978-952-219-316-2.pdf

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?	First event of the research person: birth, start of registration or migration.
2. How do individuals leave observation?	Last event of the research person: death, end of registration or migration.
3. How do households enter observation?	
4. How do households leave observation?	
5. Are some entry or exit dates unknown?	Sometimes, but then a period of exit is known
6. Are some entry or exit dates estimated?	Estimates can be made. 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?	Information is available of all individuals who were involved in a vital event (birth, marriage, death) concerning the research person.

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

1. Source(s): Which source forms the basis for the sample	There is no sample. The main sources of the database are the parish registers. All information from the end of 17th century until 1950. All parishes from the Ceded Karelia area
2. Sampling units: Households, individuals, regions…	The Karelian database is not a sample from the population in that area. The data of individuals is collecting from each church book, from every page and from every row.
3. Variables used for selection: Age, gender, marital status, other	Everybody mentioned in the parish registers of Ceded Karelia is planned to be recorded in the database.
4. Selection method: Random, stratified random, total count, clustered, other	No selection.

VIII.   Data collection

1. Data collection period: When the data was collected and transcribed?	The Karjala Database Foundation was established in 1990. Collecting the data is still going on (already over 25 years).
2. Data collection method: Public digital register, transcription, other	Manual recording from the Web pages: http://digi.narc.fi/digi/?lang=en_US
2.a. If transcription, how was the transcription done: ◦ By individuals ◦ From scanned sources ◦ From LDS's microfilms ◦ Automatic controls	Data from an Orthodox parish registers: the transcription from Russian to Finnish was done by individuals.
2.b. How was the checking of the transcription done? For example, by proof reading?	Not done yet.
2.c. When was it done?	

2.d. Purpose of the transcription: please indicate 　◦　LDS 　◦　Research 　◦　Genealogy	Genealogy
3. Control methods by researcher: e.g. Internal consistencies such as a death cannot happen before a birth of the same person	Control methods are mainly used in the recording programmes, but considering the whole database: not done yet.
4. Data collection staff: Please indicate the number of people and their position (member of the project, free-lancer, other**)**	Hired staff for the recording project.

IX.　Linkage process

1. Linkage: Which sources and units of observation have been linked: (e.g. birth/baptisms and death/burials…)?	Not done 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:*
No	Occupational titles:	HISCO developed, but not implemented
No	Locations (including geo-referenced systems):	Level of villages.
No	Religion, civil status etc.:	Civil status, sex, marks etc.: Standardized.
	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. Syntyneiden ja kastettujen luettelot - lists of Born and Baptised

1. Official name of the source and its English translation	Syntyneiden ja kastettujen luettelot - lists of Born and Baptised
2. Purpose of the source:	
2.a. Why was this source created?	The church law 1686: The law made parish registers mandatory in every parish.
2.b. Who created it?	The staff of parishes.
3. Scope: What group of the population was documented in this source?	Everybody, who was born in Karelia.
4. Time period: When the information of the sources was recorded? Please indicate the start and the end date.	From 1682 to 1950
5. Geographical area: What territory is covered by the source?	https://katiha.mamk.fi/map.jsp
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	Partially destroyed or damaged (cause of the war/churches were fired).
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 Provincial Archives of Mikkeli
9. Documentation:	
9.a. Completely documented and accessible by:	
9.b. Partially documented and accessible by:	Description of the file is accessible by the front page of Karelian Database Foundation http://www.karjalatk.fi/index.php?lang=en . Please choose Rekisteriseloste from the bottom of the page.
9.c. No documentation, but accessible by:	

The main characteristics of the source (per every type of the source)
2. Kuulutettujen ja vihittyjen luettelot - lists of Banns and Married

1. Official name of the source and its English translation	Kuulutettujen ja vihittyjen luettelot - lists of Banns and Married

2. Purpose of the source:	
2.a. Why was this source created?	The church law 1686: The law made parish registers mandatory in every parish.
2.b. Who created it?	The staff of parishes.
3. Scope: What group of the population was documented in this source?	Everybody, who was married in Karelia.
4. Time period: When the information of the sources was recorded? Please indicate the start and the end date.	From 1681 to 1950
5. Geographical area: What territory is covered by the source?	https://katiha.mamk.fi/map.jsp
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	Partially destroyed or damaged (cause of the war/churches were fired).
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 Provincial Archives of Mikkeli
9. Documentation:	
9.a. Completely documented and accessible by:	
9.b. Partially documented and accessible by:	Description of the file is accessible by the front page of Karelian Database Foundation http://www.karjalatk.fi/index.php?lang=en . Please choose Rekisteriseloste from the bottom of the page.
9.c. No documentation, but accessible by:	

The main characteristics of the source (per every type of the source)
3. Kuolleiden ja haudattujen luettelot - lists of Dead and Buried

1. Official name of the source and its English translation	Kuolleiden ja haudattujen luettelot - lists of Dead and Buried
2. Purpose of the source:	
2.a. Why was this source created?	The church law 1686: The law made parish registers mandatory in every parish.
2.b. Who created it?	The staff of parishes.
3. Scope: What group of the population was documented in this source?	Everybody who was died in Karelia.
4. Time period: When the information of the sources was recorded? Please indicate the start and the end date.	From 1688 to 1950
5. Geographical area: What territory is covered by the source?	https://katiha.mamk.fi/map.jsp
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	Partially destroyed or damaged (cause of the war/churches were fired).
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 Provincial Archives of Mikkeli
9. Documentation:	
9.a. Completely documented and accessible by:	
9.b. Partially documented and accessible by:	Description of the file is accessible by the front page of Karelian Database Foundation http://www.karjalatk.fi/index.php?lang=en . Please choose Rekisteriseloste from the bottom of the page.
9.c. No documentation, but accessible by:	

The main characteristics of the source (per every type of the source)
4.   Muutaneiden luettelot - lists of Migrated

1. Official name of the source and its English translation	Muutaneiden luettelot - lists of Migrated
2. Purpose of the source:	
2.a. Why was this source created?	The church law 1686: The law made parish registers mandatory in every parish.
2.b. Who created it?	The staff of parishes.
3. Scope: What group of the population was documented in this source?	Everybody who was migrated in Karelia.
4. Time period: When the information of the sources was recorded? Please indicate the start and the end date.	From 1847 to 1950
5. Geographical area: What territory is covered by the source?	https://katiha.mamk.fi/map.jsp
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	Partially destroyed or damaged (cause of the war/churches were fired).
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 Provincial Archives of Mikkeli
9. Documentation:	
9.a. Completely documented and accessible by:	
9.b. Partially documented and accessible by:	Description of the file is accessible by the front page of Karelian Database Foundation

	http://www.karjalatk.fi/index.php?lang=en . Please choose Rekisteriseloste from the bottom of the page.
9.c. No documentation, but accessible by:	

The main characteristics of the source (per every type of the source)
5.   Rippikirjat - Catechetical lists (Communion books)

1. Official name of the source and its English translation	Rippikirjat - Catechetical lists (Communion books)
2. Purpose of the source:	
2.a. Why was this source created?	The church law 1686: The law made parish registers mandatory in every parish.
2.b. Who created it?	The staff of parishes.
3. Scope: What group of the population was documented in this source?	Everybody in the parishes (parish members age over 15 years) in Karelia.
4. Time period: When the information of the sources was recorded? Please indicate the start and the end date.	From 1737 to 1950
5. Geographical area: What territory is covered by the source?	https://katiha.mamk.fi/map.jsp
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	Partially destroyed or damaged (cause of the war/churches were fired).
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 Provincial Archives of Mikkeli
9. Documentation:	
9.a. Completely documented and accessible by:	
9.b. Partially documented and accessible by:	Description of the file is accessible by the front page of Karelian Database Foundation http://www.karjalatk.fi/index.php?lang=en . Please choose Rekisteriseloste from the bottom of the page.
9.c. No documentation, but accessible by:	

The main characteristics of the source (per every type of the source)
6.   Lastenkirjat - Children's books

1. Official name of the source and its English translation	Lastenkirjat - Children's books
2. Purpose of the source:	
2.a. Why was this source created?	The church law 1686: The law made parish registers mandatory in every parish.
2.b. Who created it?	The staff of parishes.

3. Scope: What group of the population was documented in this source?	Every child and their families in the parishes (parish members age under 15 years) in Karelia.
4. Time period: When the information of the sources was recorded? Please indicate the start and the end date.	From 1800 to 1950
5. Geographical area: What territory is covered by the source?	https://katiha.mamk.fi/map.jsp
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	Partially destroyed or damaged (cause of the war/churches were fired).
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)?	The Provincial Archives of Mikkeli
9. Documentation:	
9.a. Completely documented and accessible by:	
9.b. Partially documented and accessible by:	Description of the file is accessible by the front page of Karelian Database Foundation http://www.karjalatk.fi/index.php?lang=en . Please choose Rekisteriseloste from the bottom of the page.
9.c. No documentation, but accessible by:	