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	National Sample of the 1901 Census of Canada
1.a. Subtitle, which brings meaning to the title (scope, place, time period):	
2. Abbreviation	

3. Links to website(s):	
3.a. Homepage	http://web.uvic.ca/hrd/cfp/
3.b. Get to data	http://web.uvic.ca/hrd/cfp/data/index.html

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 national sample of the 1901 census of Canada was created by the Canadian Families Project at the University of Victoria between 1996 and 2001. The database contains all information for sampled dwellings from Schedule 1 of the census (the population schedule) and from Schedule 2 (the property schedule). The sampling point was the dwelling number in the first column of the enumeration form. The sample was a random sample of five percent of all dwellings within each of 129 microfilm reels was selected. This means that there is a degree of geographic stratification. Information on all individual persons within each sampled dwelling was entered. The sample yielded 50,941 dwellings and 265,287 individuals. A more

	complete description of the sampling procedure and data entry methods can be found in the User's Guide. The User's Guide contains a copy of instructions to enumerators.

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.*	Census, population register, demography, history.
Please add your own keywords, if you have data not covered by the recommended terms.	Canada

6. **Citation:** Indicate how you want others to cite your database.	Canadian Families Project, National Sample of the 1901 Census of Canada.

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?	
8.a. If completed, please indicate the years of its construction?	1996-2001
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.	No future plans. This database complements the census databases created by the Canadian Century Research Infrastructure Project.

II.   Contact information

1. Name of institute or organisation	
1.a. Website	http://web.uvic.ca/hrd/cfp/
1.b. Location: city, country	History Department, University of Victoria
1.c. Postal address	PO Box 1700, Victoria BC Canada V8W 2Y2
1.d. Phone	

2. Name of primary responsible person	Eric W Sager
2.a. His/her email address	ewsager@uvic.ca
2.b. Postal address	History, University of Victoria, PO Box 1700, Victoria BC Canada V8W 2Y2
2.c. Phone	250 884 8713

3. Administrative information	
3.a. When this form was filled?	13 June 2015
3.b. Who did it?	Eric Sager

4. Main economic funding (Name of organization(s) who made the grants /sustain it)	Social Science and Humanities Research Council of Canada

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				
2.	Marriages from church registers				
3.	Burials				
4.	Population registers, maintained by church or state				
5.	Civil birth certificates				
6.	Civil marriage certificates				
7.	Civil death certificates				
8.	Population Census	Yes	1901	1901	Population Census of Canada
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	**1901**
2. Territory covered by data	**Canada**
3. Geographical characteristic: local, regional, national, cross-national	**National**

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	Yes	265,287	
2.	Married couples	Yes	50,941	
3.	Families	Yes		
4.	Households	Yes		
5.	Farms	No		
6.	Institutions	No		
7.	Other			

5. Variables per unit included in the database	
<u>On individuals:</u> *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	113 variables including constructed variables and numeric codes for occupation, religion, birthplace, relation to household head, and ethnicity.
<u>On households:</u> *Type of household, children present, age and number of children, etc.* Please add more variables, if they are not in the list	Dwellings and households, and all persons in each sampled household including kin and nonkin of household head.

6. Kinship relations:	
6.a. How is kinship recorded in the database?	See User's Guide and instructions to enumerators.
6.b. How deep (number of generations) is kinship information going?	Not deep: all living persons in each dwelling are recorded.

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	Access, SPSS

9. Access conditions:	
9.a. How does a user get access to the database?	Open access at the web site.
9.b. What are the conditions and restrictions?	None

V. Publications and reports

1. Main publications about the database itself (max. 5)

Eric W. Sager and Peter Baskerville, eds. Household Counts: Canadian Households and Families in 1901 (Toronto: University of Toronto Press, 2007).

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?	**Not applicable – database is not longitudinal.**
2. How do individuals leave observation?	
3. How do households enter observation?	
4. How do households leave observation?	
5. Are some entry or exit dates unknown?	
6. Are some entry or exit dates estimated?	
7. Can observations be linked to geographic locations?	
8. Are the dates and locations of movements within the observation area recorded?	
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?	

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

1. Source(s): Which source forms the basis for the sample	Population census of Canada, 1901
2. Sampling units: Households, individuals, regions…	Dwelling
3. Variables used for selection: Age, gender, marital status, other	
4. Selection method: Random, stratified random, total count, clustered, other	Stratified random

VIII. Data collection

1. Data collection period: When the data was collected and transcribed?	1996-1998
2. Data collection method: Public digital register, transcription, other	Keypunching from microfilm copy of census.
2.a. If transcription, how was the transcription done: ◦ By individuals ◦ From scanned sources ◦ From LDS's microfilms ◦ Automatic controls	By individual data entry staff from microfilm
2.b. How was the checking of the transcription done? For example, by proof reading?	By re-entry of a sample of records.
2.c. When was it done?	1998
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	Two consistency check programs: these are in the User Guide.
4. Data collection staff: Please indicate the number of people and their position (member of the project, free-lancer, other**)**	26 individuals under supervision of the project director (Sager).

IX. Linkage process

1. Linkage: Which sources and units of observation have been linked: (e.g. birth/baptisms and death/burials…)?	n/a
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).	Codes were created for this sample. All are in the User Gude.

Y/N/P	Reference system	Explanations:
	Occupational titles:	See User Guide
	Locations (including geo-referenced	

	systems):	
	Religion, civil status etc.:	
	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 <u>separate form</u> 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?	
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:	