

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, etc.

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	BALSAC
1.a. Subtitle , which brings meaning to the title (scope, place, time period):	Quebec population database, 1621-1971
2. Abbreviation	BALSAC
3. Links to website(s):	
3.a. Homepage	http://balsac.uqac.ca/english/
3.b. Get to data	http://balsac.uqac.ca/english/acces-aux-donnees-2/acces-fin-de-recherche/
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	<ul style="list-style-type: none">◦ <i>Scope and main goal</i> BALSAC is a database designed for the automatic construction of family histories as well as ascending and descending genealogies. The BALSAC project was initiated in 1972 at the University of Quebec at Chicoutimi and the database covers the entire population of Quebec from the onset of European settlement (at the beginning of the 17th century) up to recent years. BALSAC is used in various projects pertaining to historical and social sciences. It is also extensively

	<p>used in human genetics research (population genetics, genetic epidemiology).</p> <ul style="list-style-type: none"> ◦ <i>Time and territory covered by data</i> BALSAC covers the entire territory of Quebec, from the beginning of European settlement in the 17th century to the contemporary period. ◦ <i>Sample strategy</i> The whole population is recorded through the marriage certificates from church registers: complete for Catholic records (covering 80-90% of the population), partial for other denominations. For the Saguenay region, all the vital data (births, marriages, deaths) have been recorded. ◦ <i>Main sources</i> The information contained in the database comes primarily from marriage records.
--	---

<p>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></p>	Demography, church register, civil certificates, population genetics, history, occupations, life course, social science, migration.
<p>Please add your own keywords, if you have data not covered by the recommended terms.</p>	Genealogy, consanguinity, kinship.

<p>6. Citation: Indicate how you want others to cite your database.</p>	BALSAC database (2015). url: http://balsac.ugac.ca
--	--

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

<p>8. Has the database already been completed or it is still under construction?</p>	Completed for marriages until 1965 Under construction for births and deaths.
<p>8.a. If completed, please indicate the years of its construction?</p>	
<p>8.b. If under construction, please indicate, realized parts and when it is planned to complete it?</p>	Data entry and linkage operations of marriages are completed up to 1965 for the entire province of Quebec (for the period prior to 1800, the marriages in BALSAC were provided by the PRDH at the University of Montreal). For the Saguenay region, the database contains birth, marriage and death records of the entire population (1838 to 1971) and the linkage is completed.
<p>8.c. Please add a brief description of future plans for the database.</p>	

II. Contact information

1. Name of institute or organisation	BALSAC Project
1.a. Website	http://balsac.ugac.ca/
1.b. Location: city, country	Chicoutimi, Quebec, Canada
1.c. Postal address	BALSAC Project, Université du Québec à Chicoutimi, 555 boul. de l'Université, Chicoutimi, Quebec, Canada, G7H 2B1
1.d. Phone	(418)-545-5517

2. Name of primary responsible person	Hélène Vézina
--	---------------

2.a. His/her email address	helene.vezina@ugac.ca
2.b. Postal address	
2.c. Phone	

3. Administrative information	
3.a. When this form was filled?	February 2015
3.b. Who did it?	Jean-Sébastien Bournival

4. Main economic funding (Name of organization(s) who made the grants /sustain it)	Canadian and Quebec governments funding agencies Université du Québec à Chicoutimi Université Laval Université du Montréal McGill University
---	--

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	Y	1838	1971	For the Saguenay region.
2.	Marriages from church registers	Y	1621	1965	For the whole province of Quebec
3.	Burials	Y	1839	1971	For the Saguenay region.
4.	Population registers, maintained by church or state	N			
5.	Civil birth certificates	N			
6.	Civil marriage certificates	Y			
7.	Civil death certificates	N			
8.	Population Census	N			
9.	Nominative lists	N			
10.	Military draft records	N			
11.	Other: Genealogies	Y			Genealogies of all individuals in the database can be reconstructed back to the first immigrants to Quebec.

IV. The database: core characteristics

1. Period covered by data: give first and last year of date, if possible		From the beginning of European settlement in the 17th century to the contemporary period. 1621-1971			
2. Territory covered by data		BALSAC covers the entire province of Quebec, Canada.			
3. Geographical characteristic: local, regional, national, cross-national		Regional			
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	Y	5,000,000		
2.	Married couples	Y	2,500,000		
3.	Families	Y	2,500,000		

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	Date and place of birth and death (for Saguenay only), date and place of marriage, age, gender, marital status, religious denomination, occupation, consanguinity, place of residence.
<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	N/A

6. Kinship relations:	
6.a. How is kinship recorded in the database?	The database structure allows for the reconstruction of all individual genealogies and exact measurement of kinship relationships. Moreover, canonical consanguinity declared by the spouses in the marriage certificates is recorded.
6.b. How deep (number of generations) is kinship information going?	The database structure and depth allow for the reconstruction of kinship links up to 17-18 generations (10 generations on average). Variables are the same throughout the database.

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?	N/A

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

9. Access conditions:	
9.a. How does a user get access to the database?	No direct access to the database. Data extraction by BALSAC staff and transfer of the data to the researcher via a secure website.
9.b. What are the conditions and restrictions?	Conditions described in the "Policy on Access to BALSAC Data for Research Purposes" (available at http://balsac.ugac.ca/english/files/2011/03/DOC-01-Policy-on-Access-to-BALSAC-Data-for-Research-Purposes2.pdf). The main restriction concerns the confidentiality of the recent data (100 years or less).

V. Publications and reports

- I. Main publications about the database itself:
 1. BALSAC Annual Report 2013-2014, <http://balsac.ugac.ca/english/files/2014/12/Rapport-annuel-2014-anglais.pdf>
 2. Structure of the BALSAC database and content of the main tables
<http://balsac.ugac.ca/english/files/2012/11/I-C-245.pdf>

3. Un aperçu des méthodes de reconstitution des familles et de jumelage des données du fichier de population BALSAC, <http://balsac.ugac.ca/wp-content/uploads/2011/06/I-C-231.pdf>

II. Main or exemplary publications on research based on the database:

1. Moreau, C., Bherer, C., Vézina H., Jomphe, M., Labuda, D., Excoffier, L. (2011). Deep Human Genealogies Reveal a Selective Advantage to Be on an Expanding Wave Front, *Science*, 334 (6059), 1148-1150.
2. Vézina, H., Gauvreau, D., Gagnon, A. (2012). Socioeconomic Fertility Differentials in a Late Transition Setting: A Micro-Level Analysis of the Saguenay Region in Quebec. *Demographic Research*, 30.
3. St-Hilaire, M., Richard, L., & Marcoux, R. (2014). Individual and familial life courses in Quebec City, 1871-1911: Some considerations on two biographical data-sets. In *The dawn of Canada's century: hidden histories* (pp. 322–358). Montreal : McGill-Queen's University Press.
4. Bherer, C., Labuda, D., Roy-Gagnon, M.-H., Houde, L., Tremblay, M., Vézina, H. (2011). Admixed ancestry and stratification of Quebec regional populations. *American Journal of Physical Anthropology*, 144(3), 432–441. doi:10.1002/ajpa.21424

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, etc.

III. Observations

1. How do individuals enter observation?	Individuals enter observation through the first civil record where they appear as subject or as parent.
2. How do individuals leave observation?	Individuals leave observation through the last civil record where they appear as subject or as parent.
3. How do households enter observation?	
4. How do households leave observation?	
5. Are some entry or exit dates unknown?	All entries have dates.
6. Are some entry or exit dates estimated?	Most dates are exact, some are estimated.
7. Can observations be linked to geographic locations?	Yes
8. Are the dates and locations of movements within the observation area recorded?	No
9. Are all individuals who lived in selected households recorded? (Selection on basis of the sample or because sampled individuals are living in households)	No, there is information about individuals mentioned in the vital records but who were never present in the Quebec territory (Example: an immigrant's parents at his marriage).

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

1. Source(s): Which source forms the basis for the sample	N/A
2. Sampling units: Households, individuals, regions,...	N/A
3. Variables used for selection: Age, gender, marital status, other	N/A
4. Selection method: Random, stratified random, total count, clustered, other	N/A

V. Data collection

1. Data collection period: When the data was collected and transcribed?	Since 1972, still on-going
2. Data collection method: Public digital register, transcription, other	Original registers from parishes, microfilmed registers of parishes and digitized registers from the Direction de l'état civil du Québec.
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 	Transcription by individuals
2.b. How was the checking of the transcription done? For example, by proof reading?	1 transcription out of 10 checked by proof reading.
2.c. When was it done?	
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	Automated queries are performed regularly on the whole database in order to check for internal inconsistencies.
4. Data collection staff:	Staff hired to perform data entry (around 100 in the

Please indicate the number of people and their position (member of the project, free-lancer, other)	past 40 years)
---	----------------

VI. Linkage process

1. Linkage: Which sources and units of observation have been linked: (e.g. birth/baptisms and death/burials,...)?		Family reconstitution based on births, marriages and deaths records for the Saguenay region; on marriage records for the other regions.
2. Documentation of linking:		
2.a. Programme, manually, ...		Programmes developed at BALSAC allowing for semi-automatic and automatic linkage.
2.b. Name of software if used (and its parameters)		N/A
3. What are the rules for linking? Flags definition (list them: age, name, extra knowledge, ...)		Linkage based almost exclusively on comparison of nominative data.
4. How is each reconstructed person traceable to the original sources /transcribed data?		Each event contains a link to the corresponding source (parish record).
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 for each person in the database and for each family. A person keeps the same ID number throughout all files. In the INDIVIDU file, each person is linked to his/her parents (individual's ID, father's ID, mother's ID). In the COUPLE file, spouses are linked (husband's ID, wife's ID). Families are identified by a family number which regroups spouses and their children.
6. Linkage percentage		More than 90%
7. Quality of linkage (own evaluation)		Excellent
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).		HISCO, HISCLASS Latitude-longitude coordinates
Y/N/P	Reference system	Explanations:
Y	Occupational titles:	Nearly 63% of all occupations are coded to HISCO, HISCLASS schemes. In terms of frequency, the coded occupations cover 98% of all occurrences in the database.
Y	Locations (including geo-referenced systems):	Coded at various geographic levels and geo-referenced.
Y	Religion, civil status etc.:	Homemade dictionary for 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.

VII. The main characteristics of the sources (per every type of the source)

1. Official name of the source and its English translation	Religious records (especially Catholic marriages).
2. Purpose of the source:	
2.a. Why was this source created?	For registration of vital events
2.b. Who created it?	Religious authorities.
3. Scope: What group of the population was documented in this source?	Married individuals for the Quebec province, whole population for the Saguenay region.
4. Time period: When the information of the sources was recorded? Please indicate the start and the end date.	At the time of the event.
5. Geographical area: What territory is covered by the source?	Province of Quebec, Canada.
6. Content: What was recorded?	
7. Language of written material: original sources and documentation	Mostly in French, except for English speaking parishes.
8. Preservation and storage:	
8.a. Completely preserved	Microfilms and digital images.
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)?	Direction de l'état civil du Québec.
9. Documentation:	
9.a. Completely documented and accessible by:	
9.b. Partially documented and accessible by:	
9.c. No documentation, but accessible by:	