Building Bridges:
fusion of the BALSAC and RPQA databases
(Québec, 1621-1965)

Lisa DILLON

Hélène VÉZINA

BALSAC

UQAC
Université du Québec
à Chicoutimi

Université
de Montréal

PRDH
PROGRAMME DE RECHERCHE EN DÉMOGRAPHIE HISTORIQUE

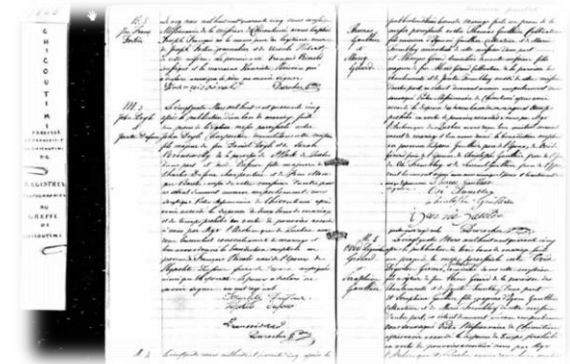**The reconstruction of colonial population from church records**
37th Annual Meeting of the Social Science History Association Vancouver, 1-4 November, 2012

Outline

- Context of historical demography in the Quebec population

- What is available now?
  Current state of the databases

- Why merge the databases?
  Rationale and motivation

- How do the databases compare?
  Some technical and « philosophical » considerations

- What form will have the "new" database?
  Our challenge
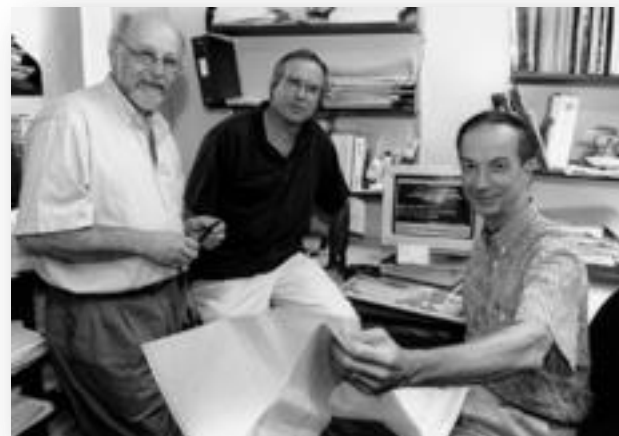
- Hopes for future developments

The Quebec context

- Availability of an exceptional documentary resource:
  the wealth of baptismal, marriage and burial certificates recorded by the Catholic Church since the early days of French settlement in the 17th century

- These certificates have been remarkably well-preserved

- « Semi-closed » population due to historical, geographical and religious reasons

- Two major population databases initiated more than 40 years ago:
  - The Registre de la population du Québec ancien (RPQA)
    Université de Montréal | Programme de recherches en démographie historique, PRDH

  - The BALSAC Project
    Université du Québec à Chicoutimi

The Registre de la population du Québec ancien (RPQA)

- Created in 1966-1967 by professors Hubert Charbonneau and Jacques Légaré (PRDH, Université de Montréal), further developped by Bertrand Desjardins

- Inspired by techniques of family reconstitution developed by the French demographer Louis Henry

- Longitudinal record linkage of the entire Catholic population of Québec from 1621-1799
(plus post-1800 deaths for persons born before 1750)

- 700 000 baptismal, marriage & burial acts 1621-1799 + 50 000 burial acts 1800-1850

The BALSAC population database

- Construction initiated in 1972 by Gérard Bouchard and his team at the Université du Québec à Chicoutimi

- First phase: family reconstitution with BMD records for the Saguenay region

- Second phase: marriage records for the province of Quebec

- Designed for automatic construction of ascending or descending genealogies (whole Quebec) and family histories (Saguenay)

Data available in BALSAC
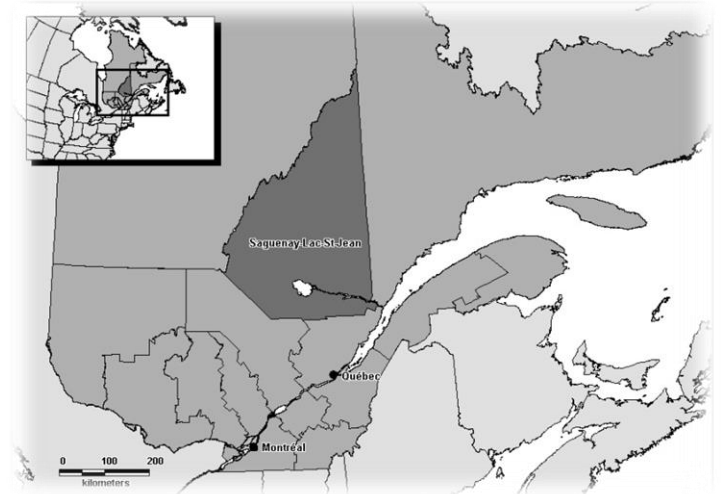
All Quebec Catholic marriages | 1621-1965

+

Births and Deaths from Saguenay | 1838-1971
(554,000 records)

=

2,15 million records
pertaining to almost **5 million**
individuals
genealogically linked

Why merge the databases? Rationale and motivation

- Preservation of a unique historical and scientific heritage: 45 years of development and operation

- Facilitate future developments

- Longer period of observation will enhance research opportunities both in social and biological fields

- Promote collaborations both at the national and international levels (comparative research)

- Maintain and develop scientific expertise and training in relevant disciplines

- Practical considerations

Linkage methodology :
the RPQA and BALSAC compared

RPQA and BALSAC

Family record based on each married couple
Record linkage based on multiple name mentions in each record
Automatic and computer assisted linkage for about 90% of cases

RPQA

- Record linkage across marriage, birth and death acts

- Names standardized via dictionary

- Manual linkage: thorough investigation of all families

- No scores assigned

BALSAC

- MOST record linkage from marriage act to marriage act (except Saguenay)

- Names standardized via phonetization

- Manual linkage: less investigation for the most recent period because of the larger number of cases

- Use of scores to indicate the confidence of links

Database Structure:
the RPQA and BALSAC compared

RPQA

- Relational Database (Microsoft SQL)

- 4 main tables: Event (Act + Mention), Individual, Couple

- Reference tables

- Names/Surnames

  - Original form matched to a standardized equivalent

BALSAC

- Relational Database (Ingres, currently moving to Microsoft SQL)

- 3 main tables: Event , Individual, Couple

- Reference tables

- Names/Surnames

  - Original and phonetized form

Our challenge…

- …an exercise in intellectual as well as methodological and technological bridge-building

- Differences in the design of the two databases rooted in their different *purposes…*

RPQA
a historical demography laboratory which...

- permits analysis of *all* demographic phenomena of a semi-closed population (prior to 1800)

- whose goal is to reconstitute the *entire* population
  - ... including persons who are observed intermittently in the parish registers (e.g. birth & marriage but no death)

- draws upon diverse acts to build an identity for each individual

- imputes identities for persons for whom certain acts are missing (or for whom certain elements within certain acts are missing)
  - .... but whose births, marriages or deaths can be inferred from mentions in other acts.

BALSAC

a *demo-genetic genealogical database* of trusted generation-to-generation links...

- ...to support demo-genetic research on the characteristics of the reproducing population (those who contribute to the genetic pool) and their ancestors

- ...to trace the depth and breadth of kinship ties of persons sharing particular genes in the present day (genetic epidemiology)

- Marriage acts feature the names of both sets of parents as well as the bride and groom, thus

  - ...BALSAC can reconstruct the entire population of all married (reproducing) persons who ever lived in Quebec on the basis of just marriage acts

Similarities which will facilitate the integration of data

- The experience of BALSAC with family reconstitution in Saguenay (BMD)

- Similarity of our linkage criteria protocols (names, permitted age gaps)

- Variables in the two databases can be harmonized

- BALSAC record linkage can be adjusted to allow for incomplete life histories

- RPQA record linkage can be adjusted to allow for the calculation of linkage scores

- Computer-assisted record linkage similar in logic

- ☞ Decision:

    1. fusion à la Intermediate Data Structure (easier, does not require harmonizing our relational database structures) or ...

    2. a full fusion of the underlying relational databases (enables simultaneous and continuous updating of the longitudinal data)

Summary of our new, fused Québec population database: # of acts

Period	Marriages	Births	Deaths
1621-1799	68 856	408 219	213 511
1800-1849	166 749	999 277	485 277
1850-1899	421 583		
1900-1949	1 040 633		
1950-1965	513 929		

690 586

1 484 554

2 142 894

1800-1849 | births & deaths from l'Institut Drouin

Summary of our new, fused Québec population database: # of cases

Period	Type of records	Region	Number of individuals
1621-1799	BMD	All of Québec	475 000
1800-1849	BMD	All of Québec	1 200 000
1850-1965	M	All of Québec	4 200 000
1850-1971	BD	Saguenay	500 000
Total			6 375 000

B = Births   M = Marriages   D = Deaths

This will allow for :
- automatic reconstruction of the genealogy of the Quebec population covering 350 years
- exploration of the historical demography of families over a 250-year period (+120 years for Saguenay)

Hopes for future developments

Integrated infrastructure of Quebec historical microdata (1621-1965): Canadian Foundation for Innovation (CFI) proposal:

- Direction de l'état civil: marriages 1966-85
- Linkage of birth and death certificates: 1800-1849
- Linkage for particular regions with 1852-1911 census microdata

The future:

- Births, marriages, deaths 1850 +
- Census microdata for all of Québec (sample, 100% indexes) 1852-1921

RPQA:
www.genealogie.umontreal.ca

Projet BALSAC:
http://balsac.uqac.ca

Mont-Valin, Saguenay
photo: Réo Blackburn