



## The Team

**Principal Investigator**  
Chad Gaffield – *UOttawa*

## Team Leaders

Carl Amrhein – *Toronto/Alberta*  
Peter Baskerville – *UVic*  
Claude Bellavance – *UQTR*  
Sean Cadigan – *MUN*  
Gordon Darroch – *YorkU*  
Charles Jones – *UToronto*  
France Normand – *UQTR*  
Evelyn Ruppert – *Trent/York*  
Eric Sager – *UVic*  
Marc St-Hilaire – *ULaval*  
Lorne Tepperman – *UToronto*  
Adam Green – *UOttawa*

## Coordinators

Carmen Bauer – *UOttawa*  
Nicola Farnworth – *YorkU*  
Byron Moldofsky – *UToronto*  
Terry Quinlan – *MUN*  
Laurent Richard – *ULaval*  
Doug Thompson – *UVic*  
Martine Tremblay – *UQTR*

## Partner Universities

- University of Victoria
- York University
- University of Toronto
- Université du Québec à Trois-Rivières
- Université Laval
- Memorial University of Newfoundland
- University of Ottawa

## Our Partners

- Statistics Canada
- Newfoundland & Labrador Statistics Agency
- Library and Archives Canada
- IBM Canada
- Institut de la statistique du Québec
- International Microdata Access Group

## INSIDE THIS ISSUE:

- Word from the Project* | 1  
*Coordinator*
- CCRI Announcement* | 1
- Contextual databases in* | 2  
*the CCRI*
- Software Innovation from* | 3  
*the CCRI*
- CCRI Sample Design* | 5
- CCRI Events* | 6

## Word from the Project Coordinator

We are pleased to present to you the fourth issue of the CCRI Newsletter. You will notice its increased volume, which is an indicator of the growing activities related to the imminent release of our databases. To date we have finished the data entry of the 1911, 1921, 1931 and 1941 censuses, the cleaning and coding of the 1921 data set is near completion, 1951 data processing has started and preparations for the release of the data bases to the RDCs are well underway.

As in the previous issues, CCRI members have taken the time to write about their work for this project. Team leaders Claude Bellavance and France Normand, and coordinator Martine Tremblay from our centre at the Université du Québec at Trois-Rivières, write about the contextual databases that will enhance the research made possible through the CCRI project.

At the heart of the project, but often enough hidden in the background, is the work of our IT team. Richard Smith, CCRI Software Application Architect, explains to us SPIDER – our in-house sampling and data entry software.

Finally, Gordon Darroch, team leader of the CCRI team at York University and chair of the Sampling Subgroup outlines the CCRI approach to sampling. I hope you will enjoy the articles and come back to us with questions and suggestions. Remember to visit our web site regularly to stay informed regarding CCRI activities.

**Carmen Bauer**  
*UOttawa*

## CCRI Announcement

Beginning in June, Evelyn Ruppert will be on academic leave from Trent University. She will be a Senior Visiting Research Scholar at The Open University's Centre for Research on Socio-Cultural Change (CRESC) (Milton Keynes, England). She will be joining her partner, Engin Isin (Prof. and CRC at York), who has taken up a Chair in Citizenship at The Open University. CRESC's objective is to develop empirically focused accounts of cultural change and its economic, social and political implications. It is a partnership with The University of Manchester and involves the disciplines of Sociology, Social Anthropology, Geography, History and our CCRI favorite.... Census and Survey Statistics! Evelyn will be working on her current pet project, *Thinking and Making Population*. Through her affiliation with CRESC she will also work on establishing collaborative links and initiatives on census-related projects with scholars in the UK.



CCRI Co-team leader, **Dr. Evelyn S. Ruppert**  
Assistant Professor, Dept. of Sociology  
Trent University (Canada)  
Visiting Senior Research Fellow, The  
Open University (UK)

*“Contextual data represent exceptionally rich material for understanding the major issues surrounding key counting events in contemporary history”*

## Contextual databases in the CCRI:

### A tool for better understanding the circumstances and issues around censuses

For the past few decades, the true value of census data has generated much debate: must the data be viewed from an objectivistic perspective or seen rather as a product, a social construct?

The CCRI team has chosen to adopt a dual approach designed to reconcile nominal lists as a source of information on a broad range of social and cultural phenomena, and the organization of censuses as a complex initiative, influenced by the major political issues of the day as well as relations between the State and civil society. Concretely, this has led us to consider, in addition to the data contained in nominal lists, a range of “data on data,” more commonly called meta-data. Contextual data form part of these meta-data. Using varied documentation, especially the media of the period, we have been able to gather valuable information on the organization, execution, and reception of censuses across Canada from 1911 to 1951.

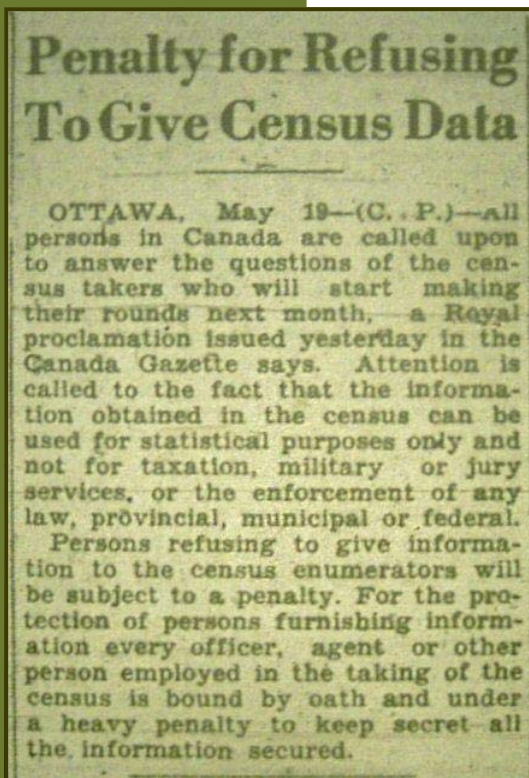
In practice, collecting contextual data appeals to a logic inverse to that of serial sources such as census nominal lists. The information is highly varied, both with respect to its form and content.

The sender and receiver of such information are key to a process-

ing strategy tailored to such types of sources. Who is talking? To whom? And on behalf of whom? Based on such questions, we see the media and institutions not simply as players but also as veritable cultural intermediaries, serving in turn as spokespersons for the government and its agencies, interest groups, and citizens.

To successfully process non-serial data, it is crucial to start out by developing a thematic processing tool. The database’s effectiveness in search mode as well as the quality and homogeneity of the information collected depend on this, with the tool serving as a reference framework for the team of assistants in charge of selecting relevant information. The tool was designed to identify the parties sending information, to situate information time-wise in relation to censuses, and to situate Canadian national censuses in relation to other forms of counting. The tool also contains the major groups of census variables.

Innovative computer solutions were required to organize the textual and iconic data. The *ContextData* database is accessed on a network using a client/server approach, with searches conducted in Web browser mode. The interface is flexible and provides tools ensuring the integrity and quality of the compiled data. This database was designed to be used independently during the CCRI’s development, after which the information will be included in the main database so that researchers who consult data from the nominal lists will be invited to learn more about the historical context in which they were generated.



*The Montreal Daily Star, 1931 May 19<sup>th</sup>, p. 33.*

In short, contextual data represent exceptionally rich material for understanding the major issues surrounding key counting events in contemporary history. They inform us on issues as fundamental as construction of the State, the mechanisms ensuring the State's legitimacy, the role of cultural intermediaries, etc. At the same time, contextual data complement other sources of "data on data" with respect to the validation of serial data drawn from nominal lists. When these data are integrated into the CCRI's central databases, they will constitute a sort of "travel invitation" for researchers interested in immersing themselves in the past.

**Claude Bellavance, France Normand  
and Martine Tremblay**  
CIEQ-UQTR

*La Presse,*  
1921 November  
12<sup>th</sup>, p. 11



## Software Innovation from the CCRI

As the quality of older historical records begins to deteriorate, archivists turn to technology to ensure that their sources are preserved for future generations. The technological solution to the preservation problem usually involves a new storage medium, which in turn has led to new tools for research. The CCRI has developed an innovative computer program, dubbed SPIDER, that fully exploits the latest form of document preservation: digital imagery.

Originally the CCRI was to obtain its 1911-1951 census data from microfilm that Statistics Canada created from census manuscripts in the 1950s. When Library and Archives Canada announced plans to digitize the 1911 census reels in 2003, the CCRI put forward a proposal to subsidize the digitization of all reels for the years within the project's scope. LAC accepted the CCRI's offer and began the conversion of the 1911 reels in September 2004.

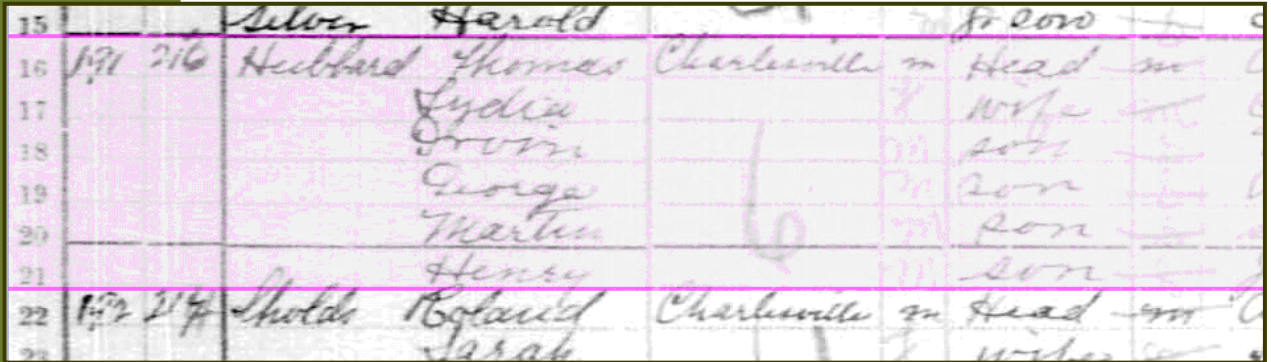
For the 1911 census, the CCRI adopted a data entry approach that allowed the team to benefit immediately from the use of digital images. A third-party computer program was used to view the digitized census manu-

scripts and a second program, developed in-house, was used in tandem for data entry. Locating each digital image was much faster than finding the equivalent information on microfilm. In addition, the viewing program can manipulate the color, sharpness and size of the images, which can greatly facilitate their interpretation.

The Sample Point Identification, Data Entry and Reporting (SPIDER) program was developed by the CCRI for the 1921-1951 censuses to build upon the advantages of the 1911 approach while eliminating the bottleneck common to most microfilm-based systems: the need for a generated number that identifies a household on a census form. This identifier, called a locator number, was required so that a household selected for inclusion in the sample by one person (the sampler) could be located by a different person (data entry operator, or DEO) who would have to capture the household data at a later date<sup>1</sup>.

<sup>1</sup>The locator number problem can be eliminated by having the person doing the sampling also capture the data as he identifies the sample points. However, there are several reasons why such an approach is not practical.

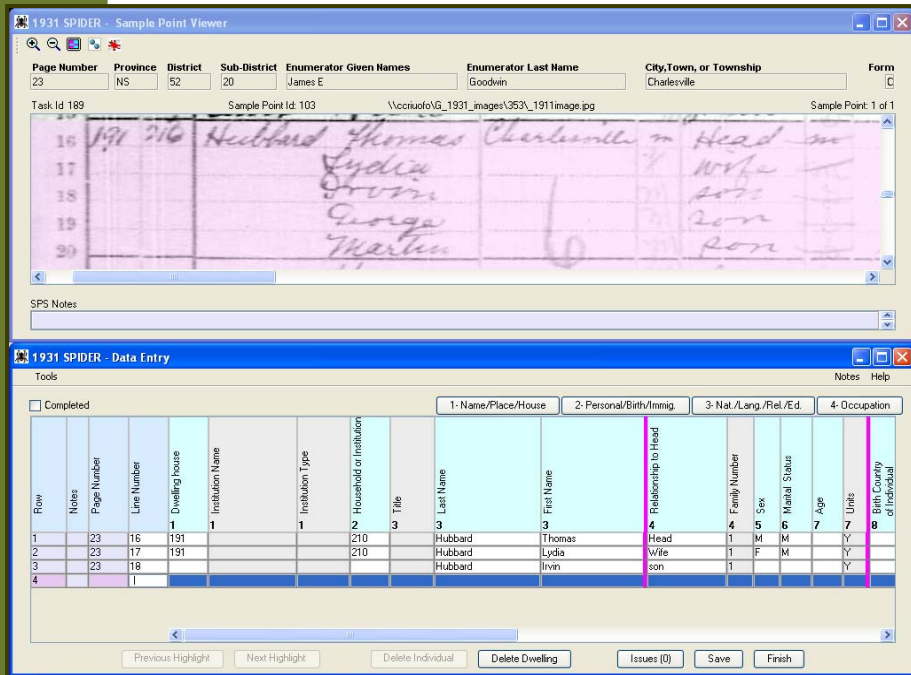
Experience has shown that transcribing and parsing the locator number is not only time consuming but is also prone to error because of the various renumbering that can appear on the census forms. SPIDER's solution exploits the potential of the medium: instead of passing a locator number between the sampler and the DEO, the software passes a digital image.



The image depicted in this article is taken from a 1911 census document, which became part of the public domain in 2003.

Figure 1 - SPIDER's highlighter is used to mark rows 16 through 21 of the image for inclusion in the sample.

The sampler uses SPIDER's image viewer component to display the digitized census form. Using the program's built-in highlighter the sampler marks the individuals on the image that must be included in the sample (figure 1). Later, SPIDER's data-capture component will present the same marked image (figure 2) to a DEO who can then quickly, and accurately, identify the data that must be captured.



The SPIDER system harnesses the potential of digital imagery. By including image attributes in its sample point definition and by integrating the selection and capture tasks, the CCRI has developed a new tool that greatly streamlines the process of capturing census data.

Rick Smith  
UOttawa

Figure 2 - SPIDER data entry. The data highlighted in the upper window must be captured in the table of the bottom window.

## CCRI Sample Designs

The sample design described here is the result of consultations among a subgroup of the CCRI team, including Professor Michael Ornstein, Director, York Institute for Social Research

The CCRI sample designs are intended to provide large, nationally representative and readily accessible samples of each of the censuses from 1911 through 1951. The censuses from 1921 to the present are strictly confidential, so the samples will be widely available only in a form that prevents the identification of individuals. More detailed analyses may be permitted in Statistics Canada's Research Data Centres. The censuses varied somewhat in the definitions of census families, households and dwellings,

but allow us to draw very comparable samples. We sample dwellings and their members. Much analysis of interest relates dwelling characteristics to those of resident families and individuals. Two parallel samples are drawn in each census year, one of "regular-sized" dwellings and a second, specialized sample of larger, "collective dwellings," most of which are institutions, such as hospitals, asylums or orphanages, or work camps. Regular-sized dwellings are defined as those with 30 or fewer residents and collective dwellings as those with 31 or more residents. Though more or less arbitrary, this cut-off size serves to divide the dwellings into those largely occupied by primary families and associated residents, such as kin or boarders from places occupied mainly by unrelated individuals.

CCRI samples are equal probability of selection (epsem) samples. They are chosen by selecting a random start within the geographic areas represented by the original microfilm reels on which the census data have been stored. Though the microdata on dwellings and individuals have been converted to digital images for our purposes, this effective geographic stratification tends to improve the national geographic representativeness of the samples as compared with a simple random sample of dwellings. Within each geographic area a systematic selection of dwellings follows the random start in a fashion that yields the desired sample densities of records. Since the Canadian population grows through these years and the proportion of the population selected is less important in analysis than the actual size of the sample, we select samples of 5 percent of dwellings (and, thus, of individuals) in 1911, 4 percent in 1921, and 3 percent in each of 1931, 1941 and 1951. This gives samples of approximately 320,000 to 420,000 in each of the 5 years.

The records for every individual within each selected regular dwelling are fully transcribed. Technically this generates "cluster samples" of individuals and of families within dwellings (since all individuals and families in every sampled dwelling are included in the sample). For some purposes, estimates of confidence intervals and standard errors should take into account such cluster effects. This can be readily accomplished with some, though not all, standard statistical software. In many analyses the samples will be so large that concern with statistical clustering will not be very important.

An unusual feature of the sample designs is how we deal with the relatively small number of "collective dwellings." The conventional solution to such a sampling issue is to create a separate sample of unusual cases in order to reduce their effects on statistical estimates of population characteristics. Our design accomplishes this through unique software that allows us to conduct a complete review of the all dwellings in the country and to make an inventory of those over size 30. We sample either separate household units or individuals within these collective dwellings in larger proportions than the main sample. This affords rare opportunities to examine institutional and related collective dwellings in detail, for example, examining the growth and changing composition of institutional populations. Analysis of these data will be strictly governed by Statistics Canada's confidentiality requirements. CCRI will provide integrated and fully documented cleaning and coding of the microdata across all the samples. This allows the new samples to be integrated with existing nineteenth and early twentieth-century samples and those that Statistics Canada, has released since 1971.

Together the two forms of samples will generate new and revised interpretations of the formative first half of Canada's twentieth century. The CCRI will also foster international, comparative analysis across increasing numbers of countries where similar databases are being constructed, and especially with the Integrated Public Use Microdata Series of the United States<sup>1</sup>.

**Gordon Darroch**  
*York University*

<sup>1</sup>See University of Minnesota Population Center, <http://www.ipums.umn.edu/>



## CANADIAN CENTURY RESEARCH INFRASTRUCTURE

Institute of Canadian Studies  
University of Ottawa  
52 University Street  
Ottawa, Ontario Canada K1N 6N5

Graphic Design and Editing

**Mirela Matiu**

*CCRI Project Assistant*

Phone: 613-562-5800 # 3657

Fax: 613-562-5216

E-mail: [mmatiu@uottawa.ca](mailto:mmatiu@uottawa.ca)

[www.canada.uottawa.ca/ccri/](http://www.canada.uottawa.ca/ccri/)



### Agence universitaire de la Francophonie

researchers whose work involves demographic dynamics and emphasizes a historical dimension. Over 100 participants from approximately 30 different countries will take part in the various workshops.

One of the sessions will discuss the Quebec and Canadian experiences and will be facilitated by Rosemary Bender, Director General, Social and Demographic Statistics, Statistics Canada. Chad Gaffield, President of the Social Sciences and Humanities Research Council of Canada (SSHRC) and Principal Investigator of the CCRI, will be the guest speaker opening this session. Marc St-Hilaire will deliver a paper, and a CCRI kiosk will provide information to participants wishing to learn more about the project. This scientific meeting is being organized by Richard Marcoux, who is the Coordinator of the AUF's *Réseau Démographie* and a full professor in *Université Laval's* Sociology Department.

<http://www.demographie.auf.org/>

## CCRI Events

**The Conference *Health in Families, Healthy Families: Gendered Explorations*** has been organized by the International Sociological Association Research Committee on Family Studies. It has been held in Toronto from May 8-11. Dr. Lorne Tepperman (CCRI Team Leader, University of Toronto) was co-organizer of this conference.

On May 9 from 11:45am to 1:00pm, there has been a session on Canadian Censuses and Family Research. This session has been organized by Dr. Peter Baskerville (CCRI Team Leader, University of Victoria), and has been chaired by Dr. Gordon Darroch (CCRI Team Leader, York University).

Presenters for this session have been:

- Dr. Chad Gaffield

Title of Paper: *The Canadian Century Research Infrastructure & Canadian Research*

- Terry Quinlan

Title of Paper: *Using Census Records to Advance Genetic Research*

- Dr. Peter Baskerville

Title of Paper: *The changing material welfare of women in urban Canada in the late 19th century*

**The 7<sup>th</sup> Journées scientifiques of the Réseau Démographie of the Agence universitaire de la francophonie (AUF)** will take place at *Université Laval* from June 19 to 22, 2007, with as theme "*Mémoires et démographie. Regards croisés au Sud et au Nord.*" The objective of this meeting will be to create a space for reflection and discussion so as to arrive at different and plural readings of social and demographic transformations. The 7<sup>th</sup> JS are for all social science professionals and

