

3 Canada

The Canadian Families Project

Peter Baskerville

Source Material

The federal Dominion government assumed responsibility for the Canadian census after confederation in 1867. The census was to be taken in 1871 and every ten years thereafter. In the 1930s the state microfilmed the nominal census enumeration forms for 1871, 1881, 1891 and 1901; microfilm copies are contained in the National Archives of Canada in Ottawa. Provincial archives also hold copies of the census for their provinces. Researchers may obtain microfilm copies of the original enumeration forms from the National Archives at about \$30 per reel.

Statistics Canada holds the censuses of 1911 and subsequent years. These sources are not currently open to researchers. Canada has no rule about access after a specific number of years; there is “returns only a guarantee of confidentiality, made early in the twentieth century. Statistics Canada has “public use” anonymized samples of 1971 and subsequent censuses, and these are available to researchers.

Some colonial-period censuses exist. Those used most

Dr. Peter Baskerville is one of the Principal Investigators of the Canadian Families Project and a member of the History faculty at the University of Victoria. His research interests include the Canadian census; the construction of family and household; and family income and standards of living. Lisa Y. Dillon of the Institute for Canadian Studies at the University of Ottawa and Eric Sager of the Canadian Families Project made valuable contributions toward the writing of this chapter.

For further information on the Canadian Families Project, see it's web site at: <http://web.uvic.ca/hrd/cfp/index.htm>.

extensively are the 1851 and 1861 censuses for the United Provinces of Canada. Industrial census returns exist for 1871 but not for subsequent years. The 1901 census includes two schedules: the nominal “returns of the living” (Schedule 1) and the schedule of “buildings and lands” (Schedule 2).¹

The Canadian Families Project (CFP) is a five-year research project funded by the Social Sciences and Humanities Research Council of Canada (SSHRC) and by five participating universities. The research team includes thirteen scholars, three of whom are postdoctoral fellows, as well as several graduate students and two computer programmers.² Disciplines represented include anthropology, demography, geography, history, and sociology. The Project has created a five percent randomized sample of the 1901 national census of Canada. The sample contains 50,943 dwellings and 265,287 individuals from a national population of 1,028,892 dwellings and 5,371,315 individuals. The Data Expansion section which follows later in this essay contains a listing and brief discussion of other computerized Canadian census files.

Procedural History

In its conception and its content the CFP database bears the imprint of its Canadian provenance in the 1990s. When the project was conceived in the early 1990s, SSHRC did not fund the creation of databases or research tools; it did (and still does) fund major collaborative and interdisciplinary *research* projects under its MCRI (Major Collaborative Research Initiative)

¹ Further information is contained in Kris Inwood and Richard Reid. (1995). “Introduction: The Use of Census Manuscript Data for Historical Research,” *Histoire sociale/Social History*, 56: 301-11. See also the other papers in this special issue on census data.

² Co-investigators are, at the University of Victoria, Peter Baskerville, Annalee Golz, Lynne Marks, Ian MacPherson, Larry McCann and Eric W. Sager; at York University, Bettina Bradbury and Gordon Darroch; at the University of Ottawa, Chad Gaffield; at the Université de Sherbrooke, Peter Gossage; at Concordia University, Danielle Gauvreau. Postdoctoral fellows are Ken Sylvester and Stacie Burke (Victoria) and Lisa Dillon (Ottawa).

program. Our national sample of the 1901 census was created, therefore, as part of an interdisciplinary research project on family in Canada. To this extent our national sample of the 1901 census differs in its origins from most public use microdata samples created from historical censuses in the United States, and from the Integrated Public Use Microdata Series (IPUMS) created by the Minnesota Historical Census Projects.

The CFP sample is a stratified random sample of five percent of all dwellings. The sampling point is the dwelling count in column 1 of Schedule 1 of the census; all persons within each sampled dwelling were entered. We counted all dwellings on each of 129 microfilm reels and entered a random sample of five percent of those dwellings. The sample is, therefore, stratified by reel. All information was entered from both Schedule 1, "Population," and Schedule 2, "Buildings and Lands, Churches and Schools." Schedule 2 gives information on property owned or leased by the individuals in Schedule 1, as well as their street addresses (for those in urban places), a variable missing from previous Canadian censuses. Table 3-1 lists all variables in the database (including a selection of the constructed variables). A sampling density of five percent, unusual for a public use sample, reflects our concern to capture sufficient cases to allow analysis at the sub-provincial level. The Canadian attention to regional differences and "limited identities" is also reflected in the number of over samples taken from the 1901 census, and stored separately from the national sample. The CFP possesses complete coverage of all individuals in fourteen communities of varying size across Canada.³

Selecting, as we have done, by dwelling, allows for a straightforward analysis of household characteristics. Households were selected because CFP researchers believe them to be critically important social units for analysis. This sample design complicates individual level analysis, however, since individuals in this sample are clustered by household and are

³ The communities are Amherst and Kentville, Nova Scotia; Campbellford and Dunnville, Ontario; Dauphin, Manitoba; Hague, Wetaskiwin, Hyde, Neudorf, Sinaluta, and Wolseley in the Territories; Nelson, Revelstoke and Victoria, British Columbia. Several graduate students have created databases from other communities for their thesis research.

not a simple random sample. Some characteristics of one household member, such as religion, can be predictors of the characteristics of other members. This is particularly a problem with small samples, somewhere below 5000 cases. For helpful suggestions on how best to control for bias when using this database for individual level analysis see Ornstein (1983).⁴

It is no longer possible, if it ever was so, to treat routinely generated information in historical sources as a transparent window into the social reality of the past. The census itself must be problematized, its provenance displayed and its internal logic, its consistencies and inconsistencies unraveled. The documentation that accompanies such databases expands and includes far more than “data dictionaries” of alphabetic strings and numeric codes. It is, therefore, deeply regrettable that almost all of the original documents of the census office within the Department of Agriculture have been lost. We know that census officials in Canada communicated with their colleagues in Britain and the United States, but our forays in the Public Records Office and the U.S. National Archives have failed to turn up any extensive correspondence. We had to begin with the instructions to enumerators and the debates in Parliament and the press over the taking of the census. Each project co-investigator then took responsibility for a specific field, or related fields within the census, and dissected those fields, seeking the patterns, the inconsistencies, the errors, and the discordant voices (see *Historical Methods*, forthcoming).

Todd Gardner, of the Minnesota Historical Census Project, adopted the software for constructing the public use sample of the 1920 U.S. census to fit the 1901 Canadian census. Marc Trottier, CFP programmer, wrote two long consistency check programs which are outlined in detail in the CFP’s *User’s Guide*.⁵ Five percent of all the 65,287 cases were proof-read by visual

⁴ Michael D. Ornstein. (1983). “Discrete Multivariate Analysis: An Example from the 1871 Canadian Census,” *Historical Methods*, 16: 101-08.

⁵ The second version of the *User’s Guide* is in preparation and will be released with the database in 2001; the first version is Eric W. Sager, D. K. Thompson and Marc Trottier. (1997). *The National Sample of the 1901 Census of Canada User’s Guide*, Version 1.0. Victoria, BC.

comparison of the data with original document: in all fields the data-entry error rate was less than half of one percent. Specific fields have been subjected to additional proof reading and correction and all French proper names have been checked against the original entry.

Variable Availability

Variables available in the 1901 public use sample are listed in Table 3-1 on pages which follow.

Confidentiality Provisions

All national censuses in Canada up to and including 1901 are available on microfilm with no restrictions on access. Censuses after 1901 and before 1961 are not available other than in published aggregate form. Public use samples in anonymized form compiled by Statistics Canada are available from 1971 to the present. There is currently a major debate in Canada concerning the accessibility of censuses after that date.

Table 3—1. Variables in the Canadian Families Project National Sample of the 1901 Census

| <u>Variable</u> | <u>Description</u> |
|---------------------------------|---|
| <u>Name</u> | |
| 1. Constructed variables | |
| DWELLID | Combines microfilm reel no., district no., sub-district letter/no., polling subdivision no., dwelling no. |
| HHID | DWELLID plus family/household no. from column 2 of Schedule 1 |
| DWPOS | Individual's position in dwelling (1st, 2nd, etc.) |
| HHPOS | Individual's position in family/household |
| 2. Schedule 1—Population | |
| HHNBR | Number of family/household |
| INDLNM | Surname of individual |
| INDFNM | First name(s) and initials |
| SEX | Sex (f or m) |
| COLOUR | Color (usually white, black, red or yellow) |

**Table 3—1. Variables in the Canadian Families Project
National Sample of the 1901 Census**
(Continued)

| <u>Variable</u> | |
|---|--|
| <u>Name</u> | <u>Description</u> |
| 2. Schedule 1—Population (Continued) | |
| RELHEAD | Relationship to head of household |
| RELHEAD2 | Numeric code for relationship to head |
| MARST | Marital status |
| BDAY | Day of birth |
| BMONTH | Month of birth |
| BYEAR | Year of birth (4 digits) |
| AGEYR | Age at last birthday |
| AGEMO | Age in months (if less than 1) |
| BPL | Country or place of birth |
| BPL2 | Numeric code for birthplace |
| URBAN | If born in Canada, is birthplace rural or urban |
| IMMYR | Year of immigration to Canada |
| NATYR | Year of naturalization |
| RACE | Racial or tribal origin |
| RACE2 | Numeric code for racial or tribal origin |
| NATL | Nationality |
| NATL2 | Numeric code for nationality |
| RELIGION | Religion |
| RELIGIO2 | Numeric code for religion |
| OCC | Profession, occupation, trade or means of living |
| OCC1 | Numeric code for Profession, occupation, trade. |
| OCC2 | Constructed variable: occupation type |
| RETIRED | R for retired |
| OWNMEANS | Living on own means |
| EMPLOYER | Employer |
| EMPLOYEE | Employee |
| OWNACCT | Working on own account |
| TRADE | Working at trade in factory or home |
| WORKPLC | Name of workplace if given by enumerator |

Table 3—1. Variables in the Canadian Families Project
National Sample of the 1901 Census
(Continued)

| <u>Variable Name</u> | <u>Description</u> |
|---|---|
| 2. Schedule 1—Population (Continued) | |
| MOEMPFAC | Months employed at trade in factory |
| EARNINGS | Earnings from occupation or trade |
| MOEMPOTH | Months employed in other than trade in factory or home |
| MOEMPHOM | Months employed at trade in home |
| EXEARN | Extra earnings from other than chief occupation |
| MOSCHOOL | Months at school in year |
| CANREAD | Can read |
| CANWRITE | Can write |
| ENGLISH | Can speak English |
| FRENCH | Can speak French |
| MTONGUE | Mother tongue |
| MTONGUE2 | Numeric code for mother tongue |
| INFIRM | Infirmities |
| INFIRM2 | Numeric code for infirmities |
| DISTRICT | District no. |
| SDISTRIC | Sub-district letter |
| POLL | Polling subdivision no. |
| DWELLING | Dwelling house no. |
| PROVINCE | Province as entered at top of Schedule 1 |
| PROVDIST | Province as constructed from District Number |
| PROV2 | Numeric code for province from |
| PROVINCE | |
| PLACE | City, town, village or township |
| ENUMFNM | Enumerator's first name |
| ENUMLNM | Enumerator's surname |
| PAGENBR | Page number from top of Schedule 1 |
| LINENBR | Line number of individual on Schedule 1 demarcated by column 1 |

Table 3—1. Variables in the Canadian Families Project
National Sample of the 1901 Census
(Continued)

| <u>Variable</u> | |
|---|---|
| <u>Name</u> | <u>Description</u> |
| 3. Schedule 1—Population (Continued) | |
| NBRINFAM | Constructed: number of persons in family/ household (as demarcated by column 2, Schedule 1) |
| NBRINDW | Constructed: no. of persons in Dwelling as |
| NBROFFAM | Constructed: no. of family/households (column 2 counts) in Dwelling |
| HINST | Name of institution if given by enumerator |
| REEL | Number of microfilm reel |
| CHKNOTE | Note entered by operator at 2nd consistency check |
| INDNOTE | Note entered by operator during data entry |
| DWNOTE | Note on dwelling by operator during data entry |
| 3. Schedule 2—Buildings and Lands | |
| PROPOWNR | Constructed: indicates (y or n) if individual has property in Schedule 2 |
| NBRPROP | Constructed: the number of lines for this individual in Schedule 2 |
| LOCATION | Place of habitation or address |
| HOUSES | House in construction |
| VACANT | Houses vacant |
| INHAB | Houses inhabited |
| BINST | Name of institution |
| BLDGS | Number of buildings |
| FAMILIES | Number of families in house or institution |
| ROOMS | Number of rooms in house or institution for each family |
| INMATES | Number of inmates in institution from |
| OACRES | Real estate owned: grand total of acres |
| OLOTS | Real estate owned: no. of town or village lots |

Table 3—1. Variables in the Canadian Families Project
National Sample of the 1901 Census
(Continued)

| <u>Variable</u> <u>Name</u> | <u>Description</u> |
|--|--|
| 3. Schedule 2—Buildings and Lands (continued) | |
| OHOUSES | Real estate owned: no. of dwelling houses |
| OSTORES | Real estate owned: no. of stores, warehouses, etc. |
| OBARNS | Real estate owned: no. of barns, stables, outbuildings |
| OSILOS | Real estate owned: no. of silos |
| OSILOCAP | Real estate owned: capacity of silos |
| OMANU | Real estate owned: no. of manufacturing establishments |
| LACRES | Real estate leased: grand total of acres |
| LLOTS | Real estate leased: no. of town or village lots |
| LHOUSES | Real estate leased: no. of dwelling houses |
| LSTORES | Real estate leased: no. of stores, warehouses etc. |
| LBARNS | Real estate leased: no. of barns, stables, outbuildings |
| LSILOS | Real estate leased: no. of silos |
| LSILOCAP | Real estate leased: capacity of silos |
| LMANU | Real estate leased: no. of manufacturing establishments |
| RELDENOM | Churches: religious denomination |
| COMMS | Churches: number of communicants |
| SEATS | Churches: seating capacity of edifice |
| SSDENOM | Sunday school: religious denomination |
| SSTCHRS | Sunday school: number of officers and teachers |
| SSSCHOL | Sunday school: number of scholars |
| PSROOMS | Public school: number of rooms |
| PSTCHRS | Public school: number of teachers |
| PSSCHOL | Public school: number of scholars |
| ENUMDAY | Day of visit by enumerator |
| ENUMMO | Month of visit by enumerator |

Data Access / Electronic Format

Since the CFP is a research project in progress and since the granting agency, SSHRC, will measure the Project's success mainly by its research output, SSHRC agreed that the CFP's national sample of the 1901 census would be reserved for the use of project co-investigators, post doctoral fellows and graduate and under graduate students until 2001. The database will be made available in April of that year in easily downloadable electronic format. Further details of public access will be available in 2001 at the CFP web site: <http://web.uvic.ca/hrd/cfp/>.

Research Possibilities

The census is our fundamental source of information about American social structure in the past. No other source can compete with respect to population coverage and reliability. For the period before the mid-twentieth century, the census provides the only data on population characteristics that are not profoundly distorted with respect to class, race, gender or education. Moreover, the census is the only historical source that provides comprehensive geographic coverage and broad chronological scope.⁶

Censuses are constructions of the national population created by individuals and groups within the state. Censuses are surveys, reporting the voices of large numbers of people, speaking on behalf of their families or households to census enumerators, answering questions framed by census officers within the Canadian Department of Agriculture. The results should be understood as a dialogue, a long series of questions and answers in which class, race, gender, language, and other influences guide the conversations.

It is a mistake even to see the questions as determined by the state or state officials alone. The questions themselves have multiple origins -- in borrowings from British and U.S. censuses of the late nineteenth century; in the desire of mainly rural-

⁶ Steven Ruggles and Russell R. Menard. (1995). "The Minnesota Historical Census Projects," *Historical Methods*, 28:6.

origin Canadian politicians to measure the “progress” of the nation and to promote further development; in the desire of politicians to understand changing language patterns in a multi-ethnic population; in the concerns of specific departments of government to understand wage labor and class conflict; and in the pressure of specific people and groups, such as the Trades and Labour Congress, for the inclusion of certain questions in the census. The number of questions in the 1901 census relating to labor force participation – no less than 14 questions! – must be understood as part of government’s response to class conflict and to the specific lobbying efforts of organized labor.⁷ The census, therefore, is a dialogue between people and state occurring at a specific moment in the history of a predominantly rural northern nation-state, containing within it more than one cultural nation, many languages, and an emerging urban working class.

The census of Canada was a *de jure* census, although this concept was not clearly defined prior to the twentieth century. The Census Acts did not define the term, but Instructions to enumerators in 1901, for instance, state that people were to be enumerated “in their home or usual place of abode.” Persons temporarily absent from home were to be enumerated at their home; only where there was no “fixed period of return” should persons be enumerated in a temporary residence. These instructions left open the possibility for considerable variation on the part of enumerators, and led to recurring public and parliamentary debate. Was the number of French Canadians being over-estimated because some enumerators in Quebec were counting people who had taken up residence in the United States? What was a “usual place of abode?” The 1901 enumerator in the Yukon enumerated all the passengers in steamboats on the Yukon River: was this their “usual place of abode?” There is a serious risk of over-enumeration, if people were enumerated in both their seasonal work location (logging camp or mining camp) and their “usual” familial residence. There is a risk of some under-enumeration where enumerators

⁷ Peter Baskerville and Eric Sager. (1995). “Finding the Workforce in the 1901 Census of Canada,” *Histoire sociale/Social History*, 56:521-44.

located transients and omitted them on the grounds that their "usual" residence was elsewhere.

The spatial boundaries of the 1901 census present serious analytical problems. Individuals were enumerated by census Districts and Sub-districts. These were meant to correspond to electoral divisions and subdivisions, since one purpose of the census was to meet the criterion of "representation by population" in the House of Commons of the federal Parliament. The 1901 census also had a smaller spatial unit – the polling subdivision; all individuals were enumerated within dwellings, numbered in sequence within each polling subdivision. Unfortunately maps of these spatial units do not exist for any historical census. The researcher is left with only a general indication of the location of each enumerated dwelling and its inhabitants (usually the name of the census district and subdistrict, and often the name of the city, town or village). Not until 1901 did the census, in Schedule 2, state the street address of the dwelling (for urban dwellings) or the rural location of the dwelling. Cartographic possibilities, and even clear distinctions between rural and urban, are severely limited prior to 1901.

A particular problem for researchers relates to the fact that enumeration was conducted in either English or French, depending on the location and linguistic mix of the population being enumerated. The understanding of dwelling, household and family varied between language groups. Particular care must be taken in reading information in specific columns. In 1901, for instance, under "Colour," the response 'b' means black where the enumerator is using the English language and white where the enumerator is using French. Lisa Dillon has determined that indices for age heaping in the 1901 census are lower for French- than English- Canadians. The tradition of vital event record keeping among parish priest in French Canadian communities enhance accuracy of age reporting among French Canadians in 1901.

Those parts of the historical censuses covering Canadian Native or aboriginal peoples present both problems and opportunities. Preliminary work suggests that under-enumeration may have been extensive. Often enumeration was done by the local Indian agent, who reported only aggregated information. Patterns of co-residence among aboriginal peoples

did not fit with the categories constructed for the rest of the population: even in 1901 – more thorough than any previous census – the numbering of dwellings is often unclear or missing altogether. Even the form used for Schedule 1 is not the same as for the rest of the country and often is incomplete; occupation, for instance, is usually missing on these forms. These returns are most useful where the enumerator had good local knowledge or was a local ethnographer, familiar with the Native languages.

It was intended that the census be taken on a specific “census day” in the spring of the census year. Clearly many enumerators were late in completing their enumeration. A systematic analysis of the actual dates of enumeration has not been undertaken, but many enumerators were still at work in July and August of the census year. In a country where much labor was seasonal and workers were highly mobile, the late enumeration could affect the results in some districts.

The enumeration went through two stages: enumerators completed their forms and sent them to the Census Office in Ottawa. At the census office collators or “correctors” ensured that forms were complete and sometimes “corrected” work done in the field. Chad Gaffield and his students have done some work on corrections to the language questions in the 1901 census. Problems appear to have been particularly serious in 1861.⁸

Systematic estimates of over- or under-enumeration do not exist for Canadian historical censuses. We know that under-enumeration of women’s labor force participation occurred. The reporting of occupation was determined by a narrow definition of labor force participation; for example, women employed on farms received no stated occupation. Even in 1901, when the possibility of multiple occupations is acknowledged for the first time, occupational pluralism remains hidden. Great care must be taken with the property information in Schedule 2 for 1901. It appears that enumerators often attributed all property within a household to the head of the household. This has consequences

⁸ See Bruce Curtis. (1994). “On the Local Construction of Statistical Knowledge: Making Up the 1861 Census of Canada,” *Journal of Historical Sociology*, 7: 416-34.

for estimates of the extent of property holding by women as well as the overall distribution of property among individuals.

In the dialogue between enumerator and respondent, answers stretched and sometimes broke the boundaries of census-takers' categories. Census takers assumed that all households could be demarcated by the presence of a male household "head". Respondents did not always share that assumption. Household headship is a highly variable concept and nearly ten percent of household heads reported were female.

Perhaps even more striking is the dialogue surrounding marital status. Canadian politicians tended to assume that there were only three possibilities: single, married, or widowed. Canada, unlike the United States, was assumed to be free of the moral contagion implied by divorce.⁹ The 1901 census confirmed the point: it was the first Canadian census to permit the entry of "d" for divorced as a marital status response. The published census states that there were only 661 divorced people in the country. But that total does not reflect what enumerators recorded! Annalee Golz and her assistants scrutinized all responses for the population of 5.3 million people on 129 microfilm reels and counted 880 divorced persons. It appears likely that officials in the Census Office "corrected" the entries, especially for Quebec, where divorce was not recognized either by the civil code or by the Catholic Church. Yet 269 Quebec Catholics stated that they were divorced. In the dialogue between people and state they made that fact known, and historians with access to the original manuscript census can hear them.

Thanks to the geographers on our team our census database contains a variable that gives the population of all communities above 1,000 population. With this variable we can now feed community size and location back into the analysis of households, family strategies, labor-force participation, church membership, and even, if we choose, single parenting, age-specific marital fertility, or property holding. A composite map of census districts has been used in the development of a geographic information

⁹ James Snell. (1991). *In The Shadow of the Law: Divorce in Canada, 1900-1939*. University of Toronto Press: Toronto.

system (GIS) for mapping spatial patterns derived from the national sample data (see Gregory and Southall, this volume). Once the map of census districts was created, these geographical units were then digitized by transferring the outline of a base map of Canada's provinces and census district boundaries to a digital form that could be used in various computer software formats. This was done using OCAD software. From this format these map files could be transferred, in the form of Digital Exchange Files (DFX), to CorelDRAW where the provinces and census districts were converted to a constant scale and pieced together to generate a computerized map of Canada. At this stage the line work was also edited to create distinct polygons (districts) for labeling purposes. The base map was then ready for use with different GIS software packages. The principal software used to date has been MAPINFO, which can create thematic choropleth maps with relative ease.

Boundary mapping of census sub-districts has been limited because of the absence of any maps accompanying the original census, and because of the enormous cost in labor involved in determining the boundaries of 3,089 census sub-districts. Census polling sub-division boundaries (the smallest of the 3 census spatial units) have, however, been reconstructed for the following cities: Halifax, Montreal, Toronto, Winnipeg, Vancouver, and Victoria.

Data Expansion

The CFP's sample of the 1901 census was never intended to stand-alone. It was always intended to begin the process of comparative analysis across space and longitudinal analysis across time. Indeed, the CFP has created oversamples (usually 100 percent coverage) of 12 small communities from the 1901 census (usually Schedule One data). In the Canadian context it is already possible to draw comparisons among several censuses.

Before the CFP created its sample of the 1901 census there was only one machine-readable national sample of a historic census, the census of 1871. Gordon Darroch and Michael Ornstein created this sample in 1979 at the Institute of Social Research at York University, Toronto, Canada. The 1871 sample is drawn from the first Canadian census, taken for the four

provinces at Confederation, Nova Scotia, New Brunswick, Quebec and Ontario. Nearly three-quarters of the population resided in southern Ontario and Quebec. The sample is a probability sample of dwellings, and thus, a *cluster* sample of individuals. Variables included in the file allow for making exact estimates of the standard errors of population statistics taking account of clustering, although standard statistical packages will not do so. All individual records within dwellings were transcribed for the first, personal schedule of the census. The 1871 sample was stratified at the first stage by province, and within provinces between urban (defined as communities of 3,000 or more) and rural areas. This stratification was intended to ensure representation of these eight groups (four provinces and urban and rural areas within each). There are 10,000 dwellings/households in the 1871 sample and just over 62,000 individual records, representing approximately 1.7 percent of the population recorded in that year. The weighted sample for making national estimates is approximately 24,700 individuals.

In a second stage several oversamples were added to increase the precision of comparisons between provinces and urban areas. These require the use of sample-specific weights when such comparisons are the object of analysis. In addition, the sampling procedure created three special subsamples to allow detailed analysis of particular ethnic populations. These include (1) an oversample of households with at least one person of German ethnic-origin or nationality, a characteristic first identified in the 1871 census, (taken from all census districts with at least 15 percent German-origin population), (2) an oversample of households with at least one person of French origin in Ontario and New Brunswick (from districts in which at least 15 percent of the population was of French origin) and (3) a sample of households in Quebec with a British-origin person, (from districts with at least 15 percent British). Appropriate weights for national estimates, and for the oversamples and subsamples, are provided as variables on the files. Documentation for the 1871 sample is included in the CFP documentation, or it can be obtained directly through the Data Archivist of the York Institute for Social Research, York University, Toronto.

Kris Inwood, of the Economics Department, University of Guelph, has created a computerized file of the Industrial Schedule for Ontario in 1871.

The Ontario Genealogical Society has created a computerized index of the heads of household for all 1871 entries for the province of Ontario [see <http://www.archives.ca/index.html>].

Since 1956 a nation wide Canadian census has been taken every five years. Five anonymized public use samples of contemporary Canadian censuses are available (1971, 1976, 1981, 1986 and 1991). Through the Data Liberation Initiative (DLI) Statistics Canada makes these public use samples available to university based researchers through data librarians at their respective university library.

The Church of Jesus Christ of Latter-day Saints (LDS) in Utah has invested approximately 200,000 hours capturing 100 percent of the 1881 Canadian census (4.3 million individuals). Currently the Institute of Canadian Studies at the University of Ottawa, in partnership with the Canadian Families Project, and the Centre interuniversitaire d'études québécoises (CIEQ, Université Laval and Université du Québec à Trois Rivières), and the LDS, has applied for money to clean and prepare the 1881 Canadian census data for general use. It is hoped that these data can be released in computerized form by December, 2002.

An even larger scale project is currently being considered for financing by the Canadian Foundation for Innovation. Headed by the Institute of Canadian Studies at the University of Ottawa under the Directorship of Chad Gaffield, and involving a large number of collaborators, including Statistics Canada, IBM, the Canadian Families Project, the CIEQ and scholars from the University of Toronto, York University, Memorial University and the University of Saskatchewan, the Canada Century Research Infrastructure Project will construct a five percent sample of the Canadian census for 1911, a four percent sample of the 1921 census and 3 percent samples for the 1931, 1941 and 1951 censuses. The result will be relational micro-data sets covering, in total, 1,829,542 cases, each including, on average, at least three dozen data points. The relational micro-data sets will be hierarchically organized and geographically coded from individual through family, dwelling and community. The

resulting datasets will support both query-tool based analysis and innovative data mining explorations including three dimensional visualization. The project is scheduled for final completion in 2007 but a preliminary version of the dataset will be released in advance of that date.

Table 3-2 describes variables that are or will be included in the public use samples for these census years. Those years in bold type have been or are in the process of being made into public use samples. These variables include those based on questions asked in the manuscript census for that year, as well as variables which will be constructed for each year.

Table 3—2. Variables Available in the Proposed 19th and 20th Century Canadian Public Use Census Files

| Description | 1871 | 1881 | 1901 | 1911 | 1921 | 1931 | 1941 | 1951 | 1961 | 1971 | 1981 | 1991 |
|--|------|------|------|------|------|------|------|------|------|------|------|------|
| Nation & Year | X | X | X | X | X | X | X | X | X | X | X | X |
| Unique Household Number | X | X | X | X | X | X | X | X | X | X | X | X |
| Household Size | (X) | (X) | (X) | (X) | (X) | (X) | (X) | (X) | (X) | (X) | (X) | (X) |
| Position in Household | X | X | X | X | X | X | X | X | X | X | X | X |
| Last Name | X | X | X | X | X | X | X | X | X | X | X | X |
| First Name | X | X | X | X | X | X | X | X | X | X | X | X |
| Surname Similarity Code | (X) | (X) | (X) | (X) | (X) | (X) | (X) | (X) | (X) | (X) | (X) | (X) |
| Sex | X | X | X | X | X | X | X | X | X | X | X | X |
| Origin / Ancestry | X | | X | X | X | X | X | X | X | X | X | X |
| Colour / Race (American sense) | X | | X | X | (X) | (X) | (X) | (X) | (X) | X | X | X |
| Age | X | X | X | X | X | X | X | X | | | | X |
| Date of Birth | | | X | X | | | | | | X | X | X |
| Place of Birth | X | X | X | X | X | X | X | X | X | X | X | X |
| Place of Birth of Parents | | | | | X | X | | | | X | | |
| Citizenship / Nationality | | X | X | X | X | X | X | X | X | X | X | X |
| Immigrant Status/ Naturalization Year | X | | X | | X | X | X | | | | | |
| Period/Year of Immigration | | | X | X | X | X | X | X | | X | X | X |

Table 3—2. Variables Available in the Proposed 19th and 20th Century Canadian Public Use Census Files (Continued)

| Description | 1871 | 1881 | 1901 | 1911 | 1921 | 1931 | 1941 | 1951 | 1961 | 1971 | 1981 | 1991 |
|---------------------------------|------|------|------|------|------|------|------|------|------|------|------|------|
| Owning or Renting | | | X | X | X | X | X | X | X | X | X | X |
| Mortgage | | | | | | | X | X | | X | X | X |
| Rental Costs | | | | | X | | X | X | | X | X | X |
| Automobile | | | | | | X | | X | | X | X | X |
| Major home appliances | | | | | | X | | | | X | X | X |
| Number of Rooms | | | | | X | X | X | X | | X | X | X |
| Period of Construction | | | | | | | | | | X | X | X |
| Type of Dwelling | | | X | X | X | X | X | X | | X | X | X |
| Value of Dwelling | | | | | | X | X | | | X | X | X |
| # of dwellings in bldg | | | | | | | X | X | | X | X | X |
| Piped running water in dwelling | | | | | | | X | X | | X | X | X |
| Bath or shower | | | | | | | X | X | | X | X | X |
| Flush toilet | | | | | | | X | X | | X | X | X |
| Occupancy length | | | | | | | X | X | | X | X | X |
| Heating Equipment | | | | | | | X | X | | X | X | X |
| Heating Fuel | | | | | | | X | X | | X | X | X |