

## World History Conference, Sydney - July 3-9, 2005



### ABSTRACT

#### Ethics, Technology and Confidential Research Data: The case of the Canadian Century Research Infrastructure Project

Chad Gaffield, University of Ottawa, Canada

This paper explores the changing social, cultural, political and technological contexts within which researchers are addressing questions of ethics and confidentiality in their research projects. The Canadian experience offers a series of complex examples of these questions beginning with the Data Liberation Initiative in the early 1990s and continuing with the Tri-Council Working Group on Ethical Guidelines for Research Involving Humans, the creation of Research Data Centres, the government's Expert Panel on Historic Census Documents, and the Canadian Century Research Infrastructure project that was begun in 2003. Based on the author's involvement with these dossiers, the paper focuses on the past fifteen years with a view toward future possibilities both for creators and users of confidential research data. Topics include legal frameworks, academic policies and disciplinary cultures, the changing definitions of privacy and confidentiality, and the financial implications of ethical positions. Special attention will be paid to technological possibilities especially as illustrated by the approach of the Canadian Century Research Infrastructure initiative.



# **Ethics, Technology, and Confidential Research Data in Canada: The case of the Canadian Century Research Infrastructure Project**

By  
Chad Gaffield

Professor of History, University Research Chair, University of Ottawa, Canada  
Principal Investigator and Project Leader, CCRI

## **I. Introduction**

During the past forty years, researchers have become increasingly interested in the systematic study of the ways in which individuals, families, households and communities have changed in keeping with new cultural, economic, social and political developments. Topics such as demographic trends, work patterns, and school attendance rates have attracted considerable attention among scholars interested in how individual lives relate to large-scale social transformations such as urbanization, industrialization and modern state formation. For such research, census enumerations have become a privileged source since they offer the potential for analysis at both the micro and macro levels. Census enumerations were often focused on individuals within their dwellings, and therefore, they hold the promise of multi-level analysis in which, for example, children can be situated within families, households, communities and countries. The ability to fulfill this promise, however, depends upon a series of interrelated preconditions related to complex issues of public policy and administrative practice as well as research funding and technological capability. What are these preconditions and what are the changing ways in which they have affected research activities in recent years?

The following discussion addresses this question by using the Canadian experience to explore the changing social, cultural, political and technological contexts within which researchers have been attempting to study census enumerations in order to gain better understandings of the making of the modern world. At the heart of this experience have been ethical issues of privacy and confidentiality. Should census enumerations be undertaken? If so, how should the enumeration be conducted? Should the results be saved? Who should be allowed to use them for research purposes? The Canadian example is illustrative of key features of the complex debate about such questions that has characterized recent decades in various countries around the world. After briefly situating this debate in its historical context and then describing recent developments such as the Data Liberation Initiative launched in 1995 and the creation of Research Data Centres beginning in 2000, this paper focuses on the strategies being used by the Canadian Century Research Infrastructure project to create micro-data samples from the 1911-1951 census enumerations (see Appendix 1). Taken in the context of other developments, these strategies suggest some of the current and future possibilities both for the creators and users of confidential research data.

## **II. The Preconditions**

The ability of scholars to use census enumerations for systematic research depends upon six preconditions that range from the obvious to the contested: 1) that a census enumeration was conducted at a particular time and place; 2) that the census is extant; 3) that microdata can and

have been created from the enumeration; 4) that the data have been properly archived; 5) that the data are accessible for research; 6) that researchers have the knowledge and competence to analyse the data appropriately. Each of these preconditions involves intricate ethical issues of privacy and confidentiality which have evolved considerably over the decades and centuries but especially in recent years. Taken together, they provide the historical framework within which the Canadian Century Research Infrastructure project is now situated.

Ideally for research purposes, each of the six preconditions would be fully met. Comprehensive census enumerations dealing with every imaginable topic would be available for all times and places in Canadian history, and would have been not only preserved but also transformed into microdata that would be made available in readily-accessible data archives along with the necessary documentation and training to enable appropriate use. The reality of the matter is quite different for many reasons beginning with the changing purposes of census enumeration and continuing through present-day debates about research access to microdata. The result is that all of the preconditions are only partially met at the moment although recent developments have been encouraging in some key respects.

### 1) Was a census enumeration conducted?

In the case of Canada, the first census enumerations were conducted in the mid-17<sup>th</sup> century and by the time that France ceded control to Britain in 1763, forty-five censuses had been completed although some were far more comprehensive than others. The British authorities did not initially use census-taking as did the French but during the early 19<sup>th</sup> century, enumerations became more frequent and the first census act was passed in 1841. The so-called modern decennial census was legislated in 1847 for the Province of Canada (which became the provinces of Ontario and Quebec with Confederation in 1867) and the first enumeration was actually undertaken in 1852, followed by 1861 and every ten years thereafter.<sup>1</sup>

In the Canadian case, therefore, the first precondition is met to the extent that there have been census enumerations conducted from the mid-17<sup>th</sup> century to the present. The character and content of these enumerations vary considerably in keeping with the different ambitions of government authorities over the decades and the changing historical context of their work. The so-called modern census begun in the mid-19<sup>th</sup> century has been remarkably similar with each enumeration including questions about individual and family identity (age, sex, ethnicity, etc.) and economic status and activity (occupation, etc.). The similarity of the census questionnaires reflects the consistent rationale for the decennial enumerations that began with the objective of allotting parliamentary seats but also included an extensive effort to enhance the government's knowledge of social and economic patterns. As stated in 1871, "a census is taken for the purpose of ascertaining, as exactly as possible, the population and resources of a country, and thereby furnishing a sufficiently correct idea of its strength and capability." In keeping with this objective, the 1871 census asked dozens and dozens of questions in a series of nine census schedules. The number of census questions increased to more than 500 in each subsequent enumeration of the later nineteenth and early twentieth centuries as greater detail was sought about the growing complexity of Canada's "population and resources." However, by 1971, a new approach was taken in which a short form (smaller number of questions) and a long form (more

---

<sup>1</sup> R.H. Coats, "Beginnings in Canadian Statistics," *Canadian Historical Review* June 1946, vol.37, no.2; and Statistics Canada, "History of the Census of Canada," [www12.statcan.ca/English/census01/info/history.cfm](http://www12.statcan.ca/English/census01/info/history.cfm).

similar to previous enumerations) were introduced to facilitate census enumerations in light of different theories about sampling and probability.

While census enumerations have been inconsistent in their content, timing, and coverage especially before the mid-19<sup>th</sup> century, the fact that so many were undertaken in Canada is worth emphasizing especially in an international comparative context; indeed, the first precondition is met better for Canadian researchers than for many of their counterparts elsewhere.

## 2) Is a census enumeration extant?

For reasons that deserve further research, not all census enumerations remain extant. For example, only the personal schedules since 1861 have survived consistently (other schedules are extant for some enumerations). The considerations that explain the loss of most census schedules before the mid-20<sup>th</sup> century are not fully known but the preliminary evidence does not point to a fulsome public policy discussion that addressed ethical issues. Rather, it appears that some schedules were simply lost while others schedules were intentionally destroyed apparently because they were considered no longer important or were considered too cumbersome to store economically despite the fact that, at the time of their creation, census officials intended that the schedules be saved for future use in studies of change over time. These officials believed that each enumeration provided a snapshot of Canadian society that could best be understood in the context of similar snapshots before and after. However, the transfer of census schedules to the Canadian national archives was not legislated and remained irregular through the 20<sup>th</sup> century. The 1871 census was transferred to the national archives in 1941 but the 1881 census did not enter archival custody until 1979. Both the 1891 and 1901 censuses were turned over the National Archivist in 1985.<sup>2</sup>

A key development occurred in the 1950s when the extant censuses were microfilmed by the national archives in the interests of long-term preservation. This practice continued through to the 1981 census under a joint project of Statistics Canada and the national archives but then was suspended half way through the microfilming of the 1991 census for lack of funding.

At the present time (Spring 2005), Bill S-18, An Act to Amend the Statistics Act, includes a provision which specifies that each census of population records will be transferred to the newly-merged Library and Archives of Canada. The Bill is awaiting second reading in the House of Commons but, given the uncertainty of the current minority government, there is widespread concern that the Bill will remain on the order paper when the government falls.

Overall, then, the population schedules of the various census enumerations in Canada are generally extant but other schedules are far less likely to have been preserved.

## 3) Can and have microdata been created from the census enumeration?

The increasing interest since the 1960s in census microdata has led to increased efforts by Statistics Canada and by the research community to create computer files for research purposes. The possibilities for such efforts are connected to the availability of the census schedules for data creation. While Hollerith cards were used for tabulations after the 1891 census (following their introduction in the United States) and mark-sense questionnaires were used in 1951, it remains unclear if any census-bureau-created machine-readable version of a census enumeration before

---

<sup>2</sup> For a recent summary of this history, see Cara Downey, "The Census of Canada from an Archival Perspective," IASSIST Quarterly Summer 2003: 5-8.

1971 is extant; in any event, it is certain that no data files for research purposes for those outside Statistics Canada were created until 1971. Beginning with this enumeration, Statistics Canada has produced PUMFs for each census, and has done so in a way to accommodate complex issues of privacy and confidentiality that have been at the heart of census enumerations and that have determined who can create microdata from the original schedules. At least since Confederation, census enumerators were issued instructions that included obligations of confidentiality. The Manual of Instructions prepared for enumerators in each census year (1871, 1881, 1891, 1901) consistently emphasized that enumerators had to keep secret the information they received from all individuals. The wording of these Instructions is quite similar for each of these enumerations. In the early twentieth century, the insistence on secrecy is repeated in the Order-in Council of 1906 and of 1911, and then included in the 1918 Statistics Act.

The *Manual of Instructions for the enumerators of the 1871 census* stated:

The enumerator will act under oath, and his duty will be to preserve the strictest secrecy, as well with respect to any verbal statements made to him as to his enumeration records. He is not permitted to show, or in any way to communicate these, to any person whatever, except to the Commissioner of his own district, or to the Chief Officer in charge thereof; both of whom also act under oath, and are forbidden, under any circumstances, to communicate anything therein contained to any person whatever, except to other sworn officers of the Department, all bound by the like prohibition.

Similarly, the *1918 Statistics Act* stated:

No individual return, and no part of an individual return, made, and no answer to any question put, for the purposes of this Act, shall, without the previous consent in writing of the person or of owner for the time being of the undertaking in relation to which the return or answer was made or given, be published, nor, except for the purposes of a prosecution under this Act, shall any person not engaged in connection with the census be permitted to see any such individual return or any such part of any individual return.

The fact that the secrecy provisions of 1918 are in keeping with and quite similar to the secrecy requirements for census information collected in 1871 suggests that there is clear continuity with respect to the relationship between the Canadian government and respondents to the census since Confederation. In each enumeration, respondents were assured that their information would not be divulged to those not officially involved in census work.

The reasons that secrecy requirements were considered necessary for each census enumeration in Canada are similar to those in other countries including a fear among some respondents that their census responses would be immediately used against them, particularly in financial ways. The Manual of Instructions for enumerators in 1871 stated that “a census is not taken for purposes of taxation, as unfortunately, many persons imagine.” Therefore, enumerators were told that: “[p]ersons having apprehensions, or showing hesitation in giving their answers, must be assured that no information they may give; and that nothing taken down in the schedules, can, by possibility, injure, or in any way affect their standing of their business.” Similarly, in 1911, for example, enumerators were told to give “positive assurance” in the course of their work “if a fear is entertained by any person that they [census information] may be used for taxation or any other object.” In the late twentieth century, this assurance remains similar;

the 1996 census form states, for example, that “[y]our personal census information cannot be given to anyone outside Statistics Canada - not the police, not another government, not another person.”

Thus, the secrecy requirements and the explanation for them have changed very little in census enumerations since Confederation. Whatever the legal framework in place at the time, census officials were obliged to guard the secrecy of enumeration information in order to prevent tax assessors and anyone else from using it against those individuals from whom it was collected.

In this context, an informal tradition developed in Canada as in various other countries in which census enumerations became public only many decades later despite the belief that research on successive enumerations would contribute to understandings of change and continuity across communities and regions. The pre-Confederation censuses were all made public by the 1930s while the 1871 enumeration was released through the archives after the 1941 enumeration. This tradition continued when the 1881 census enumeration was made public by the archives in 1979 and it was then legislated in the Privacy Act of 1983 which formalized a 92-year rule for the release of census returns. This legislation was then followed by the immediate release of the 1891 schedules and subsequently the 1901 census in 1993. It should be noted that, in the case of Newfoundland, no provision was made for long-term confidentiality and thus the all the enumerations completed before Newfoundland joined Canada are available to the public including the 1945 census.

The reasons for using 92 years as the appropriate period between an enumeration and its release reflected both the spirit of existing practice and the fact that when the Privacy Act was made law in 1983, there existed a 92-year difference between this date and the date when the 1891 Census occurred. Thus the passing of the Privacy Act with the 92-year rule ensured that the 1891 Census records could be released in 1983. At the time, it was further noted that the 92-year rule quite conveniently ensures that any census would be released to the public during the mid-point of the census cycle. The fact that the 92-year delay is situated between the 72-year period used in the United States and the 100-year rule in Great Britain was also seen to be appropriate for Canada given where the country’s traditions and values are often located by observers.

Thus, during the era of the early mainframe computers, both census officials and researchers had access to census enumerations for creating data although public access was limited to the historic returns. By the early 1970s when Statistics Canada was creating PUMFs from the 1971 census, a growing number of researchers were beginning to create microdata files from the historic enumerations. By the early 1990s, Statistics Canada had produced microdata files for the 1971-1991 enumerations while scholars working in the public domain had created a variety of microdata files for the pre-1901 period. With the launch of the Canadian Families Project in the mid-1990s, work began to create a PUMS of the 1901 census that had been released in keeping with the 92-year rule in 1993. Taken together with previous efforts, this project suggested the increasing possibility that researchers could use census microdata to study the changing character of Canada in a present-day and historical perspective.

At that point, however, Canada’s Privacy Commissioner, Bruce Phillips, began raising concerns about the legislative justification for the 92-year rule, and since the later 1990s, a debate has unfolded concerning the extent to which the Privacy Act of 1983 and the Statistics Act of 1918 are in conflict. In late 1999, the Honourable John Manley, Minister of Industry and responsible for Statistics Canada, appointed an Expert Panel on Access to Historical Census Records “to report regarding the legal, privacy and archival implications of providing [public] access to historical census record.” The members of the Expert Panel unanimously concluded

that “with the passage of time, the privacy implications of the release of the [census] information diminishes and that the passage of 92 years is sufficient to deal with such concerns.”<sup>3</sup> However, debate continued (the 1906 census was belatedly released in 2003) and while various parliamentary efforts were initiated to deal with any legal ambiguities, none have yet been successful. Bill S-18 that is currently awaiting second reading in the House of Commons does formally confirm the 92-year rule although it also calls for a consent provision for the 2006 enumeration.

Overall, then, census microdata can currently be created for the post-1906 enumerations under the Statistics Act and has been created for the 1971-2001 enumerations. Members of the public can create microdata from the extant historic census schedules through 1906, and they have done so since the 1960s for various times and places across Canada with increasing intensity thanks to funding agencies like the Social Sciences and Humanities Research Council of Canada. Unfortunately, a complete list of census microdata files has never been compiled; indeed, the significance of the third precondition of access and creation of such files can only be understood in the context of the next precondition concerning the archival preservation of research data.

#### 4) Has the census microdata file been properly archived for research use?

While many census microdata files have been created since the 1960s, their enduring value for research purposes has been undermined by the absence of any centralized facility in Canada for the proper conservation of research data files. Happily, the PUMS created by Statistics Canada for the 1971-2001 period have been properly preserved for continuing research use but there is considerable uncertainty about the whereabouts and condition of many of the files created in the public domain in past years. The absence of any research data archive in Canada is related to complex ethical questions as well as institutional and disciplinary structures and cultures involving notions of ownership and data sharing, academic recognition, and administrative priorities.

These issues were the focus of a National Research Data Archive Consultation Working Group formed in 2000 under the leadership of the National Archives of Canada and the Social Sciences and Humanities Research Council of Canada which has tried to require funded researchers to ensure the archiving of their data since the 1980s. The initial conclusions of this Working Group were indeed that “many researchers do not archive their data, and that those who do, rely either on a limited number of overworked and underfunded university data services, or on data archives abroad.”<sup>4</sup> After an initial period of intense consultative activity and considerable optimism about the possibility of support and funding, the Working Group appears to have been unable to go beyond a description of the issues and obstacles that explain the absence of a research data archive in Canada.

In this sense, the precondition that a census microdata file can and has been created is quite distinct from whether or not it can still be used for research. The general pattern in the public domain has been for researchers to create census microdata files for their own use and perhaps that of their close colleagues and students. Over time, the likelihood has been that these files are lost track of or, in the context of rapidly changing technologies, become no longer

---

<sup>3</sup> Report of the Expert Panel on Access to Historical Census Records (chaired by Richard Van Loon and composed of Chad Gaffield, Gerard La Forest, John McCamus, and Lorna Marsden), Ottawa, 2000.

<sup>4</sup> National Research Data Archive Working Group, “Phase One: Needs Assessment Report,” 2001.

useable since they are not migrated from one computer generation to the next. While some important exceptions to this pattern should be mentioned (such as the census microdata files created in the 1970s by Gordon Darroch and Michael Ornstein from the enumerations of 1861 and 1871), the absence of an effective research data archive in Canada and the disciplinary and institutional lack of support for data sharing have meant that the net result of work since the 1960s is considerably less than the sum of completed projects.

#### 5) To whom are census microdata accessible for research?

Beyond the preconditions of census microdata creation and archiving is the question of access for research purposes in the case of census enumerations conducted within restricted time-frames such as Canada's 92-year rule. As Angela Dale has recently discussed, government statistical offices "have to tread a careful balance between providing the data needed by all sections of society and maintaining the confidence of the general public who supply most of the data."<sup>5</sup> Computerization clearly offers new possibilities for achieving this balance but it also raises new challenges as software programs increase disclosure concerns especially in light of increasingly powerful data mining strategies. Since absolute safety from disclosure seems impossible, decisions about access have to take into account both the risk of disclosure and the risks associated with not permitting research that could lead to improved public policy. These decisions have become more, not less, difficult in recent years.

Rising concerns about privacy in an era of databases for credit cards and similar personal information have also led to considerable confusion and some unwarranted perceptions about research access to confidential data. The policy impact on public perceptions can be considerable; in Angela Dale's words, "a wrong belief that people can be identified in government statistics may be as damaging to public confidence as the reality – and, for many people, the two may not be distinguished." Moreover, there are financial factors related to the cost of maintaining the balance of access and confidentiality. During the 20<sup>th</sup> century, Canadian census officials responded freely as a public service to an increasing number of requests for special tabulations of individual-level census responses but, by the 1980s, government cost-recovery policies were requiring the purchase of statistical runs including special analyses of census microdata.<sup>6</sup>

The first major effort to establish a new framework for research access to microdata was the Data Liberation Initiative conceptualized by the Social Science Federation of Canada (that later joined with its sister federation to form the Humanities and Social Sciences Federation of Canada, and now named the Canadian Federation for the Humanities and Social Sciences). With the support and leadership of Statistics Canada and the Social Sciences and Humanities Research Council of Canada, the DLI was launched in 1995 and now includes all research-focused universities in Canada. The DLI gives educationally-related research access to PUMS and other data for a subscription fee that significantly enhances use by both students and professors. Rather than feeling forced to base curriculum content on freely-accessible data especially from the United States, for example, professors have begun taking advantage of DLI to develop courses on social and economic patterns using Canadian data.

---

<sup>5</sup> Angela Dale, "Research Access to Microdata: An attempt to provide a context," *IASSIST Quarterly* Winter, 2003: 16-19.

<sup>6</sup> Kirsti Nilsen, *The Impact of Information Policy* Westport, Conn., Ablex Pub. 2001.

The financial step forward of DLI was then followed in 2000 by a second major initiative to extend research access to confidential microdata. Thanks to the additional support of the newly-created Canada Foundation for Innovation, Statistics Canada collaborated in the creation of Research Data Centres where researchers outside Ottawa could gain access to data while also meeting the confidentiality requirements. The RDCs provide secure access to full master microdata files in computer laboratories that are physically and virtually protected from unauthorized use. Beginning with six universities in 2001, RDCs are now located in thirteen institutions that are supported by more than forty universities.<sup>7</sup>

The trajectory, therefore, of research access to confidential microdata such as created from post-1906 census enumerations is certainly positive since the early 1990s. The costs have been reduced and physical access has been facilitated. Nonetheless and unlike the DLI, the RDC program has generated debate within the research community especially concerning the cost of operating a centre, the uncertainty of long-term funding, and the limitations of physical access. The hope is that the progress of recent years can be continued while additional steps are also taken to further enhance research analysis especially by increased use of the internet for remote access to confidential data such as through Virtual Private Networks.

#### 6) Are researchers able to analyse the census microdata appropriately?

The final precondition for researcher use of census microdata for studies of the changing character of Canadian society involves two distinct issues: their own competencies and the availability of complementary evidence that can enable informed and appropriate analysis. The first issue has attracted considerable attention in Canada especially since the early 1990s when a National Task Force was established to examine barriers to social science research in Canada. In addition to the difficulties of research access, the task force emphasized the small number of trained researchers in Canada who were interested in, and comfortable with, the analysis of microdata. While the task force tended to blame educational institutions for this situation, the lack of trained researchers was also connected to the need for complementary evidence that would allow researchers to respond to the critique of positivistic number-crunching. While recent decades have witnessed a continuing debate often framed in the dichotomous terms of “quantitative/qualitative” or “realist/constructivist,” researchers are paying increasing attention to the importance of situating documents such as census enumerations in larger social, economic, cultural and political contexts. Rather than beginning with microdata analysis, researchers are increasingly focusing first on gaining deep understandings of how census enumerations were conceptualized, questions formulated, and responses given. Instead of viewing microdata as “objective” information about individuals, researchers now frequently start their projects by attempting to understand the intended and unintended ways in which enumerations always produce “multi-authored” documents reflecting international exchanges, government ambitions and administrative practices as well as popular cultures, local attitudes and behaviours, and a host of other features that have both overlapped and remained distinct in complex ways across time and space. In other words, the ability of researchers to analyse census microdata has

---

<sup>7</sup> Raymond F. Currie, “Research Data Centres: A Quantum Leap Forward in Social Science Research Capabilities,” *Horizons* forthcoming July 2005.

increasingly been seen in terms of access not only to the actual microdata but also to sources that enable informed and appropriate understandings of the content and process of enumeration.<sup>8</sup>

The preoccupation with the importance of “contextual data” for analysis of census microdata is an extension of the increasing interest in meta-data as a requirement for PUMS and similar research files. Although microdata have been created since the 1960s, researchers and data archivists only began to develop standards in recent years and a great deal of work remains to be done at both the level of research culture and institutional practice. Not surprisingly, therefore, strategies for the integration of contextual data and microdata remain preliminary, and further development is a top priority for the coming years if research interest and training are to be effectively enhanced across the disciplines.

### **III. The Setting, Conceptualization and Implementation of the CCRI**

The extent to which the six preconditions for research use of census microdata were met by the turn of the 20<sup>th</sup> into the 21<sup>st</sup> century defined the setting within which the CCRI was conceptualized and implemented. Between the 1960s and the recent turn of the century, progress toward the ideal fulfillment of each precondition can best be described in terms of the proverbial two-steps forward and one step back. While the commitment to continue using the census to enumerate Canadians remained strong (unlike certain other countries), the importance of archivally preserving the original schedules remained uncertain as illustrated by the current status of the 1991 enumeration. Similarly, while an increasing number of census microdata files were created, the absence of a national research data archive as well as other factors meant that some and perhaps most of those created in the public domain are no longer useable. The launching of the Data Liberation Initiative and the establishment of Research Data Centres certainly have facilitated access to research microdata but continuing concern about financial support and the need for physical presence is fueling increased interests in using computer technologies to support virtual access at reduced costs. However, increasing privacy sensitivities (often related to perception rather than practice) are ensuring that changes to research access to confidential microdata only occur after considerable deliberation and testing. Finally, the increasing demands of researchers who view census microdata in terms of multi-authored enumeration schedules have made more challenging the task of both those working under the Statistics Act and those working in the public domain.

It was in this setting that the CCRI was conceptualized as a cyberinfrastructure to enable research on the transformation of Canadian society from the late 19<sup>th</sup> to the mid-20<sup>th</sup> century. For this purpose, all six preconditions are relevant beginning with the fact that census enumerations were indeed conducted for this period. In the context of World War II, consideration was given to canceling the 1941 enumeration but, happily for researchers, the Canadian government was decided to carry on despite the pressures of the day. Secondly, the population schedules were indeed duly saved and microfilmed in the mid-20<sup>th</sup> century although the other schedules are no longer extant.

Given the 92-year rule, creation of census microdata for enumerations after 1906 can only take place within the context of the Statistics Act and thus must be undertaken by Statistics Canada. Since the focus of Statistics Canada is on present-day activities, additional funding had

---

<sup>8</sup> Chad Gaffield, “Evidence of What? Changing Answers to the Question of Historical Sources as Illustrated by Research Using the Census,” in Jeff Keshen and Sylvie Perrier, eds., *Sources, Methods, and Interdisciplinarity* Ottawa: University of Ottawa Press 2005.

to be obtained while additional staff with the appropriate historical training had to be engaged for the research and data capture required for enumerations conducted decades ago. To meet this precondition, scholars and graduate assistants from seven universities became deemed employees of Statistics Canada and took the appropriate oath of confidentiality so that they could undertake the required work to create CCRI (see Appendix 2). In addition, a Virtual Private Network was created to connect secure computer “bunkers” in the partner universities to the secure computers at the Statistics Canada head office in Ottawa. To finance this work, a proposal was submitted to the Canada Foundation for Innovation and related agencies, and approval was given so that the project could begin in 2003. Through additional partnerships with the Library and Archives of Canada as well as other institutions such as the Newfoundland and Labrador Statistics Agency, the long-term preservation of the census microdata is assured. More immediately, the Data Liberation Initiative and the Research Data Centres will offer appropriate access to researchers for different versions of the CCRI microdata in keeping with the balance between confidentiality and academic research.

The final precondition that includes the call for contextual data to enable appropriate analysis of the census microdata is being met enthusiastically by CCRI and in innovative ways that go well beyond current expectations. The CCRI will be composed of two kinds of components: databases that present the contents of historical evidence (called “primary sources”); and databases of material created by CCRI researchers to enhance appropriate analysis of the historical evidence (called “secondary sources”). We represent these components graphically as the “Green Balloons” in which each balloon depicts a type of database within the CCRI (see Appendix 3). Integral to all these databases will be a geographical framework constructed using a Geographic Information System. GIS map layers for the entire country are being created for the 1911-1951 period to enable the location, selection, aggregation and analysis of census data and related documentary and visual material.

The primary sources of the CCRI will include four major databases, each of which is depicted graphically as one of the balloons on the lower part of the image. The most important component at the heart of the CCRI will be the census microdatabases of the 1911-1951 enumerations (a future project will focus on the 1961 enumeration in order to complete coverage for the 20<sup>th</sup> century by adding CCRI to the previously-created PUMFs for the 1971-2001 period). In addition, we are building textual databases created from documents related to the administrative history of each census enumeration as well as from newspaper coverage and political debate in the House of Commons and the Senate. The goal of these textual databases is to provide researchers with the contextual evidence necessary to undertake appropriate analysis of the census microdata.

The secondary sources of the CCRI will facilitate research on the primary sources (especially the census microdata). As represented by distinct “green balloons” on the upper part of the image, these databases will range from introductions to the various census enumerations and census questions to technical discussions of data-entry and coding issues as well as bibliographies of relevant research publications. The objective of providing this material is to support research on the CCRI primary sources by helping users understand and benefit from the intensive work undertaken in order to construct the census microdatabases as well as previous and related research efforts.

## IV. Conclusion

By taking into account the ethical considerations that frame the current research use of confidential data and by exploiting new technological possibilities for virtual collaboration, the CCRI is thus building on recent advances by both the creators and users of census microdata. While debate will undoubtedly continue about the legal frameworks, academic policies and disciplinary cultures, the roles of various government institutions, and the changing perceptions and conditions of privacy and confidentiality, the ability to analyse Canadian society throughout the 20<sup>th</sup> century will emphasize in unprecedented ways the value of research on microdata. Once completed in 2008, the CCRI will enable innovative analysis of the individuals, families, households and communities that made modern Canada. Such analysis will not provide a guaranteed foundation for moving forward successfully in the 21<sup>st</sup> century but what else can we do that is more promising?



---

## CANADIAN CENTURY RESEARCH INFRASTRUCTURE PROJECT RESEARCH TEAM & PROJECT PARTNERS

---

### RESEARCH TEAM

The *CCRI* is a pan-Canadian, multi-disciplinary and multi-institutional effort to develop a set of interrelated databases centered on data from the 1911, 1921, 1931, 1941 and 1951 Canadian censuses. Administered centrally at the University of Ottawa, the *CCRI* reflects a bilingual, interdisciplinary collaboration of leading scholars, research groups, universities, and institutions from across Canada.

#### CHAD GAFFIELD

Project Team Leader, Department of History and Institute of  
Canadian Studies, University of Ottawa

#### GORDON DARROCH

Department of Sociology  
York University

#### CARL AMRHEIN

Department of Geography  
University of Toronto/University of Alberta

#### CHARLES JONES

Department of Sociology  
University of Toronto

#### PETER BASKERVILLE

Department of History  
University of Victoria

#### FRANCE NORMAND

Centre interuniversitaire d'études québécoises  
Université du Québec à Trois-Rivières

#### CLAUDE BELLAVANCE

Centre interuniversitaire d'études québécoises  
Université du Québec à Trois-Rivières

#### ERIC SAGER

Department of History  
University of Victoria

#### SEAN CADIGAN

Department of History  
Memorial University of Newfoundland

#### MARC ST-HILAIRE

Centre interuniversitaire d'études québécoises  
Université Laval

#### LORNE TEPPERMAN

Department of Sociology  
University of Toronto

#### EVELYN RUPPERT

Department of Sociology  
York University/Trent University

### PARTNER INSTITUTIONS

- ❖ The Newfoundland and Labrador Statistics Agency
- ❖ The International Microdata Access Group (IMAG)
- ❖ L'Institut de la statistique de Québec
- ❖ Statistics Canada
- ❖ The Library and Archives of Canada
- ❖ IBM Canada



TEAM PHOTO, DECEMBER 2003

The *CCRI* is supported by the Canada Foundation for Innovation, Statistics Canada, the Library and Archives of Canada, the Ontario Innovation Trust, the Harold Crabtree Foundation, IBM Canada, le ministère de l'Éducation du Québec (formerly le Fonds FCAR), and the British Columbia Knowledge Development Fund, among other partners. The *CCRI* gratefully acknowledges their support.

---

<http://www.canada.uottawa.ca/ccri/>

---

**CANADIAN CENTURY RESEARCH INFRASTRUCTURE**  
**INFRASTRUCTURE DE RECHERCHE SUR LE CANADA AU 20<sup>È</sup> SIÈCLE**



Structure administrative / Organizational Structure – Spring 2005

