

Conceptualizing and Constructing the Canadian Century Research Infrastructure

CHAD GAFFIELD
CCRI Principal Investigator
Department of History
University of Ottawa

Abstract. The Canadian Century Research Infrastructure (CCRI) is an interdisciplinary, multi-institutional, and internationally linked initiative to enable research on the making of modern Canada. At the heart of the CCRI are microdatabases centered on the manuscript census enumerations for 1911, 1921, 1931, 1941, and 1951. This research infrastructure will be added to the results of other projects that cover the periods from 1852 to 1901 and to the Statistics Canada (STC) census microdatabases from 1971 to 2001. When completed in 2008, the CCRI will thus enable research to be made on the individuals, families, households, and communities that experienced the complex transformations of Canada since the mid-nineteenth century. By analyzing approaches to the epistemological issues involved in building the CCRI, the author seeks to advance scholarly debate by describing the research infrastructure's distinguishing characteristics and explaining its various components that seek to both support and facilitate research projects. This overview provides the context for the three other articles in this theme issue of *Historical Methods* that focus on CCRI's sampling and census microdata management strategies as well as the initiative's georeferencing and contextual data systems.

Keywords: Canadian censuses (twentieth-century), census confidentiality constraints, geographic information systems (GIS), interdisciplinarity

The distinguishing characteristics of the Canadian Century Research Infrastructure (CCRI) reflect both the pioneering achievements of previous decades and the current scholarly debate and institutional structures. Moreover, the CCRI is confronting the challenge of building a virtual research infrastructure in the rapidly changing cyber environment.

The following discussion lays out our approaches to the epistemological issues involved in building the CCRI by first describing the distinguishing characteristics of the research infrastructure and then explaining its various components, which are designed to support and facilitate

both current and future research projects. These approaches were developed in the context of seven university partners as well as team leaders across Canada, including Peter Baskerville and Eric Sager (University of Victoria); Carl Amrhein, Charles Jones, and Lorne Tepperman (University of Toronto); Gordon Darroch and Evelyn Ruppert (York University); Claude Bellavance and France Normand (Université du Québec à Trois-Rivières); Marc St-Hilaire (Université Laval); Sean Cadigan (Memorial University of Newfoundland); Adam Green (University of Ottawa); and Chad Gaffield, who serves as principal investigator at the lead institution, the University of Ottawa.

The point of departure for our work was the concept of *research infrastructure*, one that has generally been associated with the natural sciences, engineering, and biomedical research but is now becoming a central feature of scholarly activity in the social sciences and humanities (e.g., Hacker and Fitch 2003). The concept assumes that the material will be used by other individuals as well as its creators for both ongoing and not-yet anticipated research efforts. In the Canadian context, the general tradition in the human sciences of special-purpose research in which scholars create research tools to address their own questions has been particularly significant, because research data archives are limited to relatively small operations in certain universities. Although considerable discussion has occurred in recent years concerning the creation of national research data archives for Canada, it is still not easy for most scholars to deposit their research data for long-term preservation and access other than by benefiting from the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan, or the Minnesota Population Center at the University of Minnesota. Moreover, funding agencies for the social sciences and humanities in Canada

have traditionally supported the creation of research material only for the purposes of addressing the specific research questions posed by applicants for funding. Thus, research activities in Canada tend to be linked to the interests and ownership of the scholars who undertake them and are not seen as larger contributions to the research community.

The establishment of the Canada Foundation for Innovation (CFI) in 1997 heralded a new era in support for scholarly work by focusing exclusively on grants for research infrastructure. Although initially targeting only the natural sciences, engineering, and biomedical fields, the CFI was soon convinced that all disciplines needed and deserved support for research infrastructure, which came to be defined as “state-of-the-art equipment, buildings, laboratories, and databases required to conduct research.”¹ The CFI normally funds up to 40 percent of a project’s infrastructure costs, so that applicants must also seek support from other eligible institutions and partners. Fortunately, Canada’s provincial governments characteristically match the CFI support, thereby challenging the researchers to find the remaining 20 percent. It was in this context that the CCRI leveraged seed money from universities to achieve full funding in 2003 from CFI, with matching provincial grants from the Ontario Innovation Trust, Le Ministère de l’Éducation du Québec, and the British Columbia Knowledge Development Fund. In addition, we received significant support from our principal partner, Statistics Canada (STC), as well as IBM, the Newfoundland and Labrador Statistics Agency, the Harold Crabtree Foundation, L’Institut de la Statistique du Québec, and other partners, including Library and Archives Canada.

Although CCRI was funded as a research infrastructure, the proposal’s success in peer-review adjudication depended on the articulation of a compelling research agenda that would justify its construction, as well as point to the probability of continued use in light of subsequent yet unidentified research questions. Our focus emerged from previous studies of both the nineteenth and twentieth centuries that raised fundamental questions about the making of modern Canada. The studies had emphasized that, two decades after the passage of the British North America Act of 1867, Canada was still a sparsely populated, quasi-colony with a highly dependent economy and settlements spread across a vast territory on the international periphery. A century later, in contrast, Canada was a highly urbanized and industrialized country on the world stage. How did this happen? What are the social, economic, demographic, cultural, and political changes that explain Canada’s profound transformations during the twentieth century? To what extent have these changes varied across the land and among different individuals and groups? And how do these changes in Canada compare with, and relate to, developments in other countries around the world?

To address these complex questions, we decided to focus on census enumerations and a vast array of related

evidence.² In the case of Canada, the first census enumerations were conducted in the mid-seventeenth century during the time of New France, and by the time that France ceded control to Britain in 1763, 45 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 nineteenth 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 (Coats 1946).³

The character and content of these modern enumerations have been remarkably similar, with each enumeration including questions about individual and family identity, as well as economic status and activity. 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. This rationale is both related to and different from those employed in Great Britain and the United States, the two countries with which officials in Canada kept in close contact. As stated for the 1871 enumeration: “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. In 1971, a new approach was taken in which a short form (fewer questions) and a long form (more closely resembling previous enumerations) were introduced to facilitate census enumerations in light of different theories about sampling and probability. For reasons that deserve further research, not all census schedules remain extant for each enumeration, although unlike certain other countries (e.g., New Zealand), at least the personal schedules have been quite consistently preserved for all Canadian censuses (Olssen and Hickey 2005).

Although the CCRI team leaders had undertaken previous research projects using the manuscript schedules, the choice of the census as the key source for building the CCRI required considerable justification in Canada, given the trajectory of scholarly debate since the time of the early census-based research projects in the 1960s and 1970s. From the “new social history” to the cultural history that followed the “linguistic turn” during the 1980s and 1990s, scholars have focused on census enumerations for quite dif-

ferent reasons and in quite different ways. At the core of this changing research interest have been different answers to the question of evidence: census enumerations are evidence of what?

During the initial period of research activity, scholars in Canada turned to the census for two key reasons: to study evidence of behavior and to learn about the historically anonymous. Under the leadership of Michael Katz (1975), Marcel Trudel (1973), David Gagan (1981), and other researchers, the census came to be seen as a way to move beyond the literary sources that characteristically underpinned the established historical scholarship. These sources were criticized as only providing “impressionistic” evidence of the thoughts, ambitions, and claims of a small number of leaders, both official and unofficial. The census promised to enable interpretations based not on the ideas of a minority, but rather on the behavior of the whole society. Partly as a rejection of the “history of ideas” dominant among historians by the 1950s, the new compelling questions in historical debate during the 1960s and 1970s concerned what was being done across populations rather than what was being written by elites. By the early 1980s, however, scholars in Canada as elsewhere expressed an increasing skepticism about the ability of the manuscript census to provide reliable evidence about mass behavior. The initial tendency to take at face value the responses to the various census questions written on the enumeration forms gave way to growing doubt that these forms deserved to be studied at all. Using a vocabulary of errors, inaccuracies, pitfalls, bias, and limitations, researchers increasingly documented the difficulties of carrying out a census in settings such as nineteenth-century Canada. Scholarly debate moved from early skirmishes among researchers about the usefulness of specific census questions to full-blown battles about the evidentiary value of any enumeration for understanding the anonymous, specific communities or social and economic transformations (see, for example, Brookes 1976 and Fortier 1984).

Surprisingly, perhaps, given the rapidly changing scholarly orientation toward a new cultural history, the census once again became a focus of historical research in Canada by the early 1990s. However, by then, the census was defined as providing evidence of elite perspectives, values, and ambitions. In this view, it was the census questions and not the answers that were of historical interest. Census enumerations were examined not as offering ways to reveal the hidden histories of communities, but rather as evidence of state projects in which those in leadership positions attempted to enhance their power over their jurisdictions both domestically and imperially (Anderson 1991; Porter 1986; Hacking 1990). Rather than depicting enumerators writing down the information provided by householders in answer to questions of general importance, scholars now described how census officials imposed elite conceptions on the rest of society by designing both the questions and

acceptable answers. Whereas researchers in the 1960s and 1970s had focused on the “taking” of the census, their counterparts in the 1980s and 1990s wrote about the “making” of the census (Curtis 1994, 2001; Dunae 1998). Each enumeration was studied not as evidence of behavior, but rather as evidence of elite political, economic, cultural, and social ideas. And rather than counting responses to census questions, researchers studied the questions themselves and the enumeration process as illustrations of the ways in which governments were attempting to increase and solidify their power. In this way, the census came to be associated with “qualitative” research, rather than the increasingly disparaged “quantitative” research developed during the earlier period of the new social history.⁴

Even though some scholars remained attached to epistemologies associated exclusively with quantitative or qualitative, social, or cultural approaches, other researchers began attempting to move beyond these dichotomies by emphasizing that the appropriate analysis of census evidence depends on understanding the political, social, economic, and cultural contexts within which enumerations took place. From this perspective, the analysis of the responses written on enumeration schedules presupposes the study of other evidence that contextualizes these responses in their historical settings.⁵ By the mid-1990s, new efforts were made to examine various enumerations that provided evidence of individual lives that were both linguistically constructed and materially based. The census was now seen as enabling insights about diverse perceptions and realities relating to specific communities, as well as larger jurisdictions (Darroch and Soltow 1994 provided an example of an early effort in this direction).

As more scholars agreed that it was important to analyze the enumeration process and census questions related to elite priorities and perceptions, they also increasingly emphasized that census officials could not simply ask whatever they liked; enumerations depended, at least to some extent, on a shared grid of understanding about both the questions and the expected answers. Such understanding was not always present, of course, but rather than viewing the contradictory and confused results in these cases as indications of the inadequacy of the census as a historical source, scholars began seizing on such results as opportunities to probe the mentalities of those being enumerated. In the same way, researchers tended to move away from the idea that the census enumerations provided evidence only of the ability of those in power to impose their concepts and to define individuals and groups according to their own preferences. Rather, the census suggested a sense of inadequacy and ignorance among officials as well as a sense of comfort and control (Baskerville and Sager 1995, 1998). Moreover, scholars increasingly argued that ideas and behavior had to be studied in holistic ways, for they were inherently interrelated in the histories of individuals

and groups. Researchers now perceived deep connections between the histories of the anonymous and the famous with differential flows of influence both up and down, depending on the historical context, including the uneven distributions of power and influence (Gaffield 1991). In other words, the convictions of both the new social history and the subsequent cultural history seem warranted.

In keeping with this spirit, the CCRI is intentionally designed to build on the strengths of both social history and cultural history by enabling an integrated sociocultural epistemology. Our approach is based on the concept of multiauthored sources: Each source is created in direct and indirect ways by numerous authors ranging from senior census officials, middle-management census officials, correctors, census enumerators, and respondents, as well as business, religious, and community leaders both in Canada and elsewhere. From this perspective, the appropriate analysis of the census depends on understanding the multi-layered political, social, economic, and cultural contexts within which enumerations took place (Gaffield 2005). In addition to creating census microdata sets using the questions and answers written on enumeration schedules, the CCRI thus offers to researchers both documentary evidence and analysis of the cultural, geographic, and ideological character of the changing census questions and responses for each enumeration. Instead of viewing these questions as neutral inquiries and their responses as benign "facts" ready for immediate analysis, the CCRI attends to the historical context within which the census questions were constructed and answers provided. Defined as "contextual data," the CCRI provides users with historical evidence related to the substantive content of the census microdata under examination. These contextual data can be conceptualized as helping provide the "blueprints" not only for the appropriate construction of CCRI but also for successful analysis building by researchers in the coming years.⁶

In justifying a focus on census enumerations because they offer the opportunity to integrate the strengths of both the social-history approach of the 1960s and 1970s and the cultural-history emphasis of the 1980s and 1990s, the CCRI initiative is building upon the success of the Canadian Families Project (CFP), based at the University of Victoria, which created a 5 percent sample of the 1901 census enumeration under the leadership of Eric Sager and Peter Baskerville.⁷ Although the focus of this effort was the creation of census microdata, the project also inspired considerable research on the historical context of the 1901 enumeration as well as analysis of the linguistically constructed character of significant aspects of the census (Sager and Baskerville 2000, 2007). Similarly, the CCRI is building on the recent achievements in geographical information systems (GIS) of the CFP and other initiatives, especially in the United Kingdom and the United States. With mapping activity centered at Université Laval for the geographic units of Quebec and Atlantic Canada, and at the University

of Toronto for those in Ontario, western Canada, and the North, georeferencing is undertaken at the census subdivision level. This approach exceeds the robustness of our sample sizes, but it anticipates subsequent enhancement of the CCRI in light of the rapidly expanding research field interested in the creation of analytic units of geography such as a space called "coastal communities." In addition, the geographic units are enriched by published census data to enable additional multilevel analysis.

One major difference between this initiative and previous efforts is that the CCRI is focused on census enumerations whose confidentiality is protected by the Statistics Act, as well as other legislation. During the course of the twentieth century, an informal tradition developed in Canada, as in various other countries, in which each census enumeration became public only many decades later. The preconfederation censuses taken before 1867 were all made public by the 1930s, whereas the 1871 enumeration was released through the national archives after the 1941 enumeration. This tradition continued when the 1881 census enumeration was made public by the archives in 1979; it was then legislated in the Privacy Act of 1983, which formalized the 92-year rule for the release of census returns.⁸ It should be noted that, in the case of Newfoundland, no provision was made for long-term confidentiality, and thus all the enumerations completed before Newfoundland joined Canada in 1949 are available to the public, including the 1945 census.

Given the 92-year rule, the creation of census microdata for the enumerations included in CCRI can take place only within the context of the Statistics Act and thus must be undertaken by STC staff in their role for the responsible government agency. For this purpose, all CCRI team members, including scholars and graduate student assistants, undergo security checks (including the taking of the appropriate oath of confidentiality) and become "deemed employees" of STC. In addition, a virtual private network has been created to connect physically and virtually secure computer "bunkers" in the partner universities to the similarly secure computers at STC in Ottawa.

In light of the assumption that understanding Canada must emerge from a deep understanding of the country's diverse provinces and territories, the first phase of our work entailed establishing CCRI centers in the universities and institutions of our team leaders from St. John's to Victoria (see fig. 1). The CCRI center for Atlantic Canada is located at the Newfoundland and Labrador Statistics Agency in partnership with Memorial University of Newfoundland. For Quebec, the Centre interuniversitaire d'études québécoises has two CCRI centers located at Université Laval and Université du Québec à Trois-Rivières. The CCRI center at York University and the University of Toronto are covering Ontario, while the University of Victoria is hosting the CCRI center responsible for western and northern Canada. By following Canada-wide CCRI policies, practices, and standards, each center undertakes data entry as well as the construction of other

CANADIAN CENTURY RESEARCH INFRASTRUCTURE INFRASTRUCTURE DE RECHERCHE SUR LE CANADA AU 20e SIECLE

Structure administrative/Organizational Structure—Spring 2007

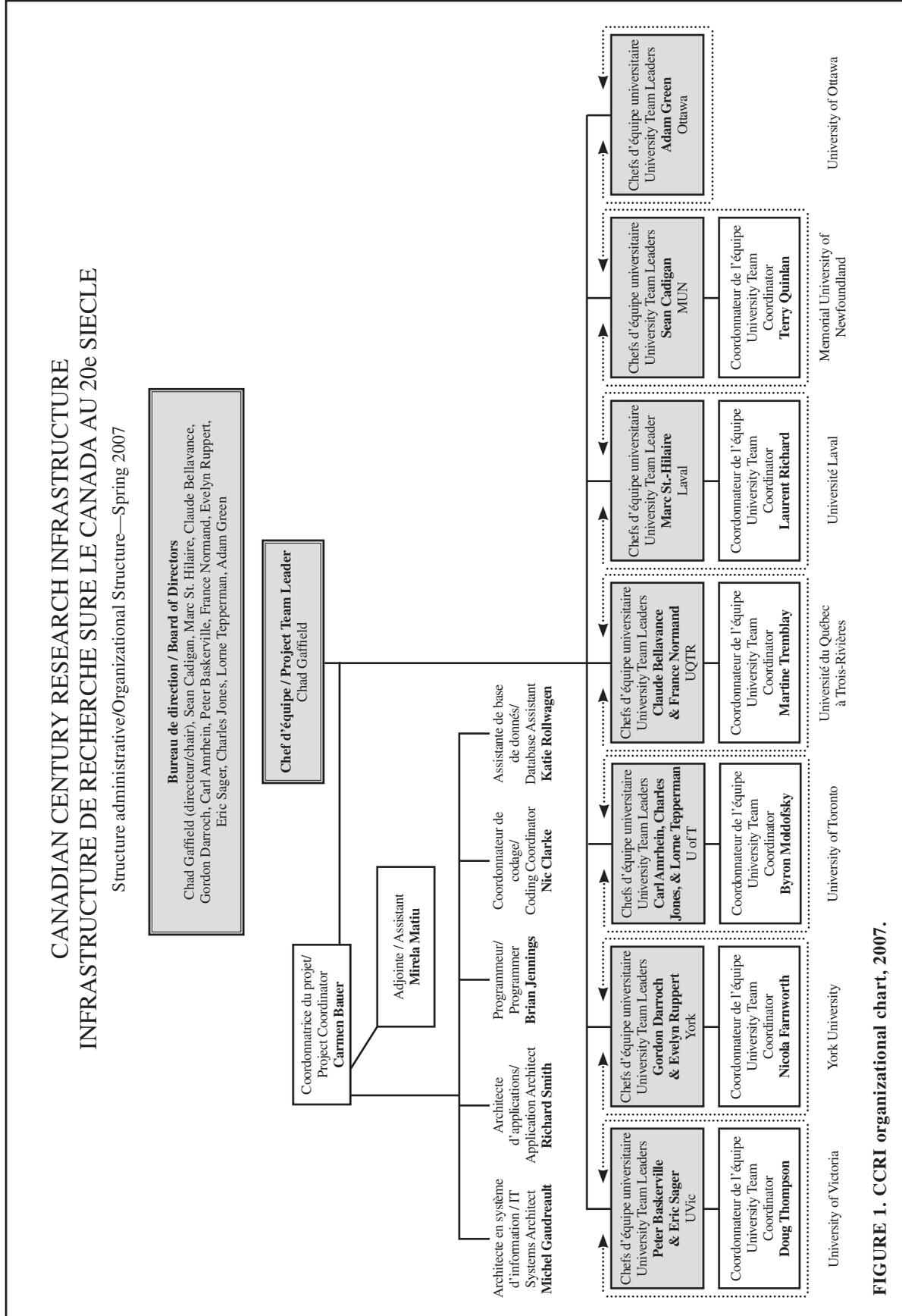


FIGURE 1. CCRI organizational chart, 2007.

CCRI components including, for example, data creation from documentary sources relevant to their part of Canada. The overall coordinating institution for the CCRI is the University of Ottawa, where the head office of the project is located at the Institute of Canadian Studies. This organizational and institutional structure assumes that Canada is certainly more than the sum of its parts, but it clearly acknowledges that both Canadian and comparative interpretations must be based on a research infrastructure created with sophisticated knowledge of different communities from sea to sea to sea. The value of this approach is illustrated by the expert local knowledge that made possible the georeferencing of more than 3,000 place-names in Newfoundland and Labrador, where numerous small and often remote coastal communities have appeared and disappeared over the decades or changed names in ways that are poorly documented.

In keeping with the inclusive character of the CCRI team, the project operates in both French and English, Canada's two official languages. It is surprising perhaps that this approach is quite unusual in Canada and contrasts markedly with the tradition of scholarly solitude in which distinct anglophone and francophone research cultures developed during the twentieth century. The CCRI language policy calls for passive bilingualism, in which team members can express themselves in the official language of their choice but are expected to understand both languages. Only selected internal documents are translated. However, given the ambition of pan-Canadian and international comparative research use, the CCRI will be made publicly available with both French and English options.

In addition to building bridges between Canada's official-language scholarly communities, the CCRI team is also practicing interdisciplinarity through a collaborative effort involving historians, geographers, and sociologists. This collaboration goes beyond the familiar exchange of concepts and methods in anticipation of interdisciplinary encounters with the completed CCRI. In other words, the CCRI is being constructed in ways that accommodate the disciplinary expectations and priorities of all team members. Rather than choosing a singular way to capture data or process evidence, the CCRI includes multiple strategies with a view toward enhancing use by diverse researchers. Practicing interdisciplinarity by involving representatives of different disciplines in building a research infrastructure is labor intensive but unavoidable if the ambition is to fulfill diverse scholarly expectations. This approach has required team leaders to be involved in every stage of construction to ensure that their priorities and disciplinary cultures are built into the various aspects of the CCRI. The result has certainly made the building of the CCRI more complex, but the hope is that this effort will be rewarded by increased use not only across disciplines but also through interdisciplinary encounters in which interpretations are debated meaningfully, for they are not focused on concepts or methods but rather involve confrontation with a common corpus of evidence.

The distinct expertise of each CCRI team member contributes to construction activities through participation in subgroups in which specialists focus on specific topics and make decisions about substantive issues related to georeferencing, sampling, and data capture as well as contextual data and groups of census questions, such as those relating to identity (e.g., religion and language) and livelihood (e.g., occupation and earnings). The CCRI subgroups normally include at least one member from each center across the country; thus, this approach connects the seven university nodes of the CCRI collaborative structure "horizontally." Whereas specific policies and decisions are left in the hands of the subgroups as much as possible, the overall coordination of their work is the responsibility of the CCRI Ottawa office where the information technology (IT) subgroup also interacts with the others to ensure proper implementation of these policies and decisions.

The CCRI team benefits from computer-based support in unprecedented ways. The project's hardware and software were acquired through a partnership with IBM that began with recognition of the vast potential for computer-based research in the social sciences and humanities. An innovative feature of this partnership is the use of DB2, IBM's high-end relational database software that can accommodate the diverse types of CCRI data, as well as support data warehousing and software such as Infominer for data mining research. Although this software currently supports all aspects of the CCRI census microdata creation and processing, certain previously begun components of the research infrastructure use other programs, such as FileMaker Pro for the newspaper research that includes digitized images, as well as extensive annotations. In addition to research-based software, team leaders and project coordinators also use VoIP telephones that not only reduce long-distance costs but also offer other features such as instant on-demand conference calls. This system began as a pilot project with Mitel, the IP communications company, and it has proven to be a key part of our communications strategy that underpins our centrally coordinated decentralized structure.

Taken together, therefore, the CCRI includes eight distinguishing characteristics: (1) a sociocultural epistemological underpinning; (2) a provision for spatial analysis through GIS; (3) microdata situated within a pan-Canadian framework; (4) partnerships with STC, Library and Archives Canada, the Newfoundland and Labrador Statistics Agency, and other agencies; (5) multiuniversity collaboration; (6) bilingualism; (7) the practice of interdisciplinarity; and (8) innovative use of IT through IBM and Mitel.

These characteristics express themselves in the actual content of the research infrastructure as depicted in the form of the CCRI balloons (see fig. 2). This image shows that the CCRI is composed of interrelated components of a relational database that includes two types of data: (1) historical evidence (*primary sources*) and (2)

material written by scholars to enhance appropriate research on the historical evidence (*secondary sources*). All components of the CCRI are georeferenced at an appropriate level of specificity ranging from specific census subdivisions to larger spatial units. The central feature of the CCRI is the census microdata files from the 1911–1951 enumerations involving individual-level stratified cluster samples ranging from 5 percent to 3 percent of the total population. This component includes five files (one for each enumeration between 1911 and 1951) composed of the responses to the census questions on the personal schedules (approximately three dozen questions for each enumeration), as well as the values for additional variables including those derived by CCRI researchers from census questions (e.g., household structure) and those created by census officials on the enumeration schedules (e.g., changed responses or added material such as occupational codes).

The intellectual justification for the inclusion of each item in the various CCRI components is made on the basis of their relationship to a specific census question (or group of questions) or to an overall census enumeration. This relationship has been defined as either *core*, meaning that the item is directly related to a census enumeration, or as *supplementary*, meaning that the item is relevant in an indirect way to the topics of a census enumeration. The CCRI is aiming to be comprehensive with respect to the core primary and secondary sources, whereas it is expected that the supplementary material will continue to be added for the foreseeable future as researchers identify more and more ways to enhance the contextual features of the CCRI. The fact that the CCRI is a cyber infrastructure will facilitate this continuing expansion.

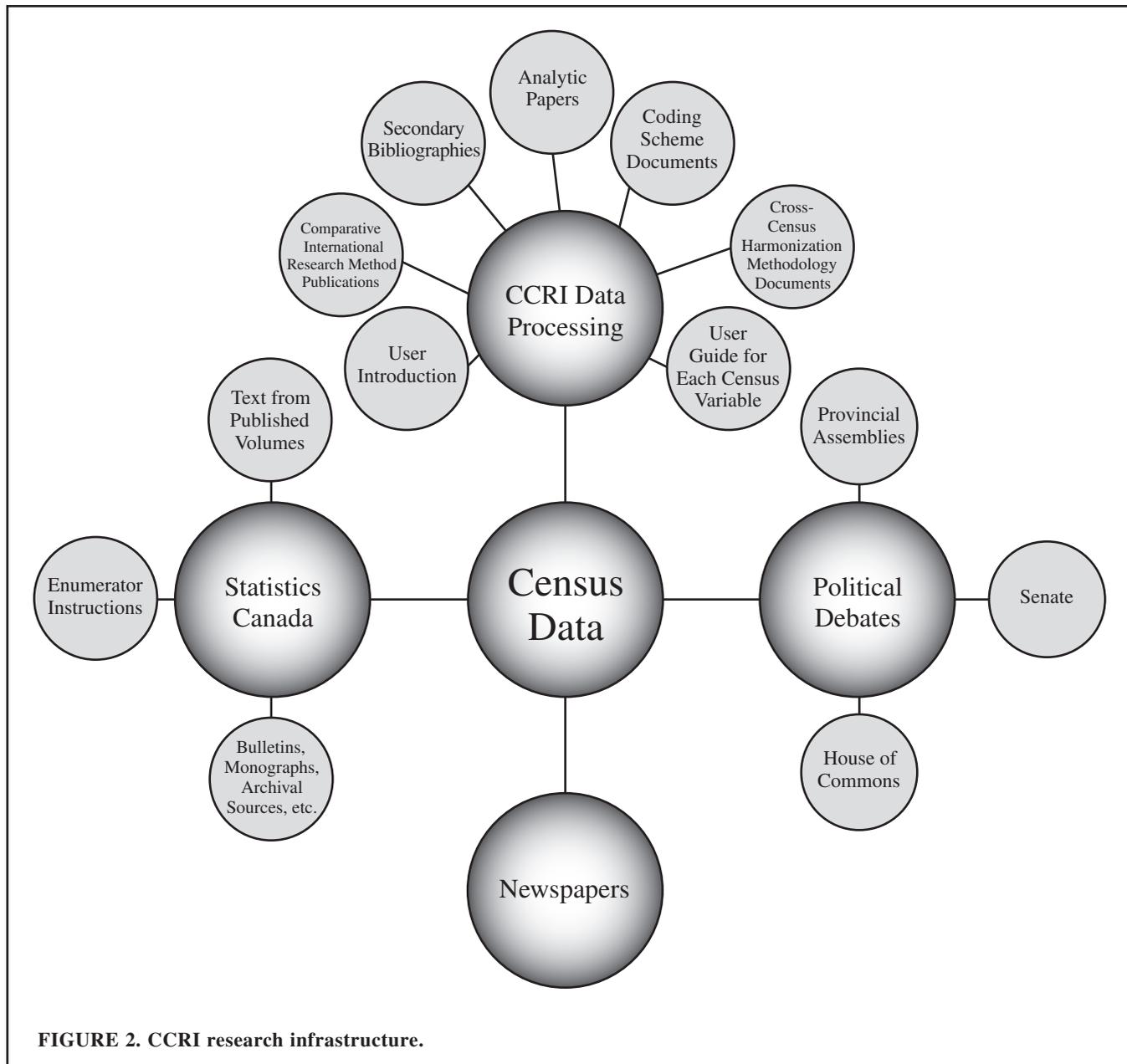
The first type of CCRI component—the primary sources—includes four major sections as depicted by the lower half of figure 2. In addition to the census microdata from the 1911 to 1951 enumerations, the CCRI offers researchers three types of evidence concerning the administrative history of each enumeration. The most familiar feature of this history involves the enumerator instructions to officers, commissioners, and enumerators for each enumeration. In addition to presenting the text of these instructions, the CCRI has also enhanced the value of this material by systematically interrelating the various parts of the instructions, so that researchers can appreciate the full implications of their specifications. To help researchers situate the enumerator instructions within the larger administration of the census, the CCRI offers two significant documentary corpuses of official publications and of manuscript documents that provide evidence of the “multiple authors” whose contact with the census office has left traces in archival holdings as well as in published materials. The published material includes text from printed census volumes (1911–1951), census monographs containing officially commissioned studies and analyses (1921–1951), census bulletins (1921–1951), STC annual reports (1918–1951), and administrative reports (1921–1951). The manuscript material is diverse, ranging from technical

memoranda to correspondence and special reports. Only part of this material had ever previously been made accessible or even known to researchers, and, in fact, its richness far surpasses our own expectations.

In addition to offering researchers primary sources created as part of the administrative history of the census enumerations, the CCRI also offers researchers evidence about the contemporary context of these enumerations. The opportunity to undertake research on this context is intended to enhance understandings of the concepts, definitions, and larger objectives associated with each census enumeration, and the ways in which these understandings were contested and debated in communities across Canada. The two types of historical evidence selected for this purpose are formal political debate and newspaper coverage. Using official reports of parliamentary proceedings, this component offers users verbatim records of what was said (or written in reports) in the federal House of Commons and the Senate as well as in the provincial legislative assemblies about any aspect of census enumeration. Thus far, our data creation for political evidence has focused on the federal level, and the extent to which the provincial sources will prove to be as valuable is not yet clear, although our expectation is that some will be richer than others in keeping with the diversity of the provinces. In contrast, we are now convinced of the consistent importance of evidence about the census in newspapers throughout the first half of the twentieth century. Using a newspaper article as the unit of presentation, the CCRI provides the original text of newspaper coverage relevant to and appropriate for the analysis of specific census questions as well as the enumeration process.

Represented on the upper half of figure 2, the secondary sources included in the CCRI are designed to facilitate research in several ways, ranging from introductory descriptions of the census enumeration process to technical discussions of data entry and coding issues, to bibliographies of census-research publications. The objective of this component is to facilitate research on the CCRI primary sources by providing users with descriptions and analysis of significant features of the census data, the administrative census history, the newspaper evidence, and political debates. For this purpose, CCRI data processors undertake analysis at three levels: (1) at the enumeration level, (2) at the question level, and (3) at the response level. For each level, the CCRI informs users about topics such as under- and overenumeration, analyses of the concepts, definitions, and objectives of each question, and studies of the meaning of various responses and how they relate to the development and implementation of coding schemes.

Specifically, the CCRI provides an overall user guide for each census year describing when and how that enumeration was conducted and noting special features that may affect the interpretation of its results. Then, one user guide for each census variable is offered to enhance



appropriate analysis of that variable. An additional CCRI component describes how each variable was coded, including the rationale for each coding scheme. In keeping with the CCRI's spirit of inclusiveness, certain variables (e.g., occupation) are coded in multiple ways according to several different approaches, and thus each coding scheme is discussed to help researchers undertake analysis according to their own preferences. In addition to user guides for each enumeration and for each question, CCRI data processing also includes analytic papers that offer detailed discussion of specific topics considered especially noteworthy or relevant to the usefulness and user-friendliness of the research infrastructure. For example, an analytic paper is being written about the changing discussion and policies regarding the

term *Canadian* as an officially unacceptable but frequently used response to the origin question. In addition to the actual census microdata, this analytic paper is based on documents from the administrative history of the census as well as from newspaper coverage and political debate. CCRI team leaders select the topics to be treated in such papers on the basis of suggestions that arise during the various data-processing activities. It is anticipated that the number of such papers will increase steadily as other researchers begin undertaking their own analyses of the CCRI.

In addition to documents prepared by data processors to enhance the research value of the CCRI, a bibliography of secondary sources lists journal articles, books, and other publications as well as conference papers and other reports

of research activities related to the Canadian censuses. Similarly, a secondary bibliography of relevant international research aims both to situate the Canadian enumerations in an international context and to reveal the links between census work in Canada and similar activities in other countries, especially the United States and Great Britain, as well as in the United Nations and other international settings. Items selected for inclusion in this bibliography provide users with a comparative perspective or methodological considerations that may affect studies of the CCRI.

Because the overall structure of the CCRI is itself composed of five complete *layers* of primary and secondary sources—each layer corresponding to one of the enumerations between 1911 and 1951—the CCRI also includes cross-census harmonization bridges or *crosswalks* that connect each of the decennially focused layers to enable comparative analysis. These crosswalks describe continuities and changes related to the questions, definitions, and other features of each census enumeration; they alert researchers to methodological considerations when responses to census questions between years are compared. In addition to such discussions, these crosswalks offer to researchers, whenever possible, strategies for undertaking comparative research. For example, a crosswalk will offer strategies for taking account of the fact that two different age groups were targeted for a specific census question in two different enumerations.

In anticipation of considerable interest in comparative analysis over time and space, the CCRI initiative collaborates with other important projects both in Canada and internationally. The CCRI's key international partner is the Minnesota Population Center (MPC) under the direction of Steven Ruggles at the University of Minnesota. The MPC's innovative efforts include the Integrated Public Use Microdata Series (IPUMS) that offers samples from 15 U.S. censuses, and the North Atlantic Population Project (NAPP) that brings together microdatabases from Canada, Great Britain, Norway, Iceland, and the United States as a way to support comparative analysis of historical change.⁹ Domestically, the 1881 Canadian Census Project began as a collaboration of the Institute of Canadian Studies (University of Ottawa) and the Church of Jesus Christ of Latter-day Saints and is now being continued at the Université de Montréal under Lisa Dillon's leadership. This project has produced a database of the complete 1881 census containing records for over 4.3 million residents of Canada. Dillon is now extending this effort to include the 1852 enumeration, with support from the CFI.¹⁰ Similarly, the CFI has made possible the 1891 Canadian Census Project under the leadership of Kris Inwood and Doug McCalla at the University of Guelph.¹¹

In developing the various components of the CCRI, we are attempting not only to build on the significant conceptual and technological achievements of previous initiatives but also to contribute to related efforts in Canada as well as internationally. We send a copy of our monthly report to

those involved in historic census-related projects and publish a CCRI newsletter to update an international network of interested individuals and groups on our construction progress. These strategies are designed to ensure that our work contributes to the efforts of others and increases the value of the CCRI for researchers both in Canada and elsewhere. Such efforts are especially important because we cannot take for granted that the CCRI microdatabases will attract substantial research attention across the social sciences and humanities in Canada. Indeed, a national task force was established in the early 1990s to examine barriers to social science research in Canada. In addition to the difficulties of preservation and access to research data, the task force emphasized the small number of trained researchers in Canada who were interested in, and comfortable with, the analysis of microdata. Although 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 widespread critique of positivistic number crunching.

The fact that a continuing debate often framed in the dichotomous terms of *quantitative/qualitative* or *realist/constructivist* helps explain why the CCRI aims to attract those researchers who are paying increasing attention to the importance of situating documents (e.g., census enumerations) in larger social, economic, cultural, and political contexts. In this sense, we expect that some researchers will begin their analysis with the microdata, whereas others will first focus on gaining a deep understanding of how census enumerations were conceptualized, questions were formulated, and responses were given. In fact, our expectation is that some researchers may plan to study only the CCRI's contextual data for projects on international exchange, government ambitions, and administrative practices as well as popular cultures and local attitudes. However, the option to move on to examine the census microdata will always be available, and we anticipate that some researchers who never expected to undertake quantitative research will end up accessing the microdata to pursue questions raised in the contextual data. Through our partnerships with STC and Library and Archives Canada, as well as other institutions (e.g., the Newfoundland and Labrador Statistics Agency), the long-term preservation of the CCRI is assured. In the meantime, the Data Liberation Initiative and the research data centers will offer appropriate access to researchers for different versions of the CCRI microdata in keeping with the balance between confidentiality and academic research requirements on census enumerations conducted within restricted time frames, such as Canada's 92-year rule. 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. Absolute safety from disclosure seems impossible, so 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 (Nilsen 2001).

The expectation, then, is that the CCRI will support diverse interdisciplinary research projects by historians, sociologists, demographers, geographers, and economists, as well as literary scholars and those in cultural studies. In addition, specific aspects of the CCRI will attract the attention of specialists who have never before studied census enumerations: for example, media scholars will be able to contribute to our understanding of the history of communications by focusing on the question posed in the 1931 census that asked if each household had a radio. Researchers will be able to examine aspects of World War II by studying the relevant questions about military service posed in the 1941 enumeration. Urban and rural planners will be able to inquire about housing, size of dwellings, and place of residence by comparing responses to the housing questions asked in 1911 and 1921. And, in all cases, scholars will be able to study these census microdata along with substantial primary sources presented as "contextual data." These examples as well as those that arise in the following articles on specific aspects of our work are only some of the ways that the CCRI will support unprecedented analyses of Canada's transformative decades in comparative and international settings. But even more important, the CCRI will raise questions about the making of twentieth-century Canada that we have not yet even imagined.

NOTES

I am very grateful to the CCRI team members and anonymous readers for their helpful comments on earlier versions of this article.

1. The CFI's mandate is to strengthen the capacity of Canadian universities, colleges, research hospitals, and nonprofit research institutions to carry out world-class research and technology development of benefit to Canadians; see <http://www.innovation.ca>.

2. The research *problématique* builds on an earlier major effort that produced the *Historical Atlas of Canada*, a three-volume work that offers an unprecedented visual depiction of time and space from 10,000 BC to the mid-twentieth century.

3. The history of the Canadian census is available on the STC Web site: <http://www12.stacan.ca/English/census01/info/history.cfm>.

4. For discussion of research activities through the mid-1980s, see José Igartua (1988). This special theme issue includes texts by Gérard Bouchard, Hubert Charboneau, David Gagan, Gordon Darroch, and Chad Gaffield. For examples of work between the mid-1980s and mid-1990s, see Kris Inwood and Richard Reid (1995). For comparative perspectives, see Robert Swierenga (1990) and David Kertzer and Dominique Arel (2002).

5. Indeed, it was record-linkage projects that, in the 1960s, first pointed to the need to contextualize census data because researchers found that, to an unanticipated extent, different sources did not cumulatively or consistently support the same interpretations of individual or collective lives. The most important scholar working in Canada on the epistemological and philosophical aspects of this topic was Ian Winchester (1970), who collaborated with Michael Katz at the Ontario Institute for Studies in Education.

6. The preoccupation with the importance of "contextual data" for analysis of census microdata is also related to the increasing interest in metadata as a requirement for research infrastructures. 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. Under the guidance of Ernie Boyko, Ann Green, Wendy Watkins, and Wendy Thomas, the CCRI is following the path of the Data Documentation Initiative (Block and Thomas 2003).

7. The Canadian Families Project 1901 sample is available on the Web site of the Data Liberation Initiative; unfortunately, it is available through Canadian universities only (see <http://web.uvic.ca/hrd/cfp/> and <http://www.statcan.ca/english/Dli/contents.htm>). Happily, a CD of the 1901 sample can be purchased at cost by contacting the principal investigator, Eric Sager, Dept. of History, University of Victoria.

8. Concerns raised by Canada's privacy commissioner delayed the release of the 1906 and 1911 enumerations. In June 2005, after a lengthy controversy, the passage of Bill S-18 (An Act to Amend the Statistics Act) approved a respondent option for preservation of the 2006 and 2011 enumerations but confirmed that each census of population records through 2001 should be transferred after 92 years to the newly merged Library and Archives Canada (Van Loon et al. 2000).

9. The North Atlantic Population Project partners seek to maximize the harmonization and usefulness of electronic census data for understanding the North Atlantic world as it entered the industrial age by offering Web-based access to microdata within a consistent format; see <http://www.nappdata.org>. See also Patricia Kelly Hall, Robert McCaa, and Gunnar Thorvaldsen (2000).

10. Gordon Darroch and Michael Ornstein created the first national sample of a Canadian census, for 1871. It has been available since the early 1980s from the archivist at the York Institute for Social Research (Oram@yorku.ca). This effort produced the first public use national historical sample of census microdata made available to researchers. For a progress report on the 1852 census project, contact Lisa Dillon at ly.dillon@umontreal.ca; for the 1881 Canadian Census Project, see <http://prdh3.demon.umontreal.ca/dillon/1881/1881.html>.

11. For information about the Historical Census Committee at the University of Guelph, see <http://www.uoguelph.ca/history/census/index.htm>.

REFERENCES

- Anderson, B. 1991. *Imagined communities: Reflections on the origins and spread of nationalism*. Rev. ed. London: Verso.
- Baskerville, P. A., and E. W. Sager. 1995. Finding the work force in the 1901 census of Canada. *Histoire sociale/Social History* 28 (56): 522–39.
- . 1998. The census and historical sources. In *Unwilling idlers: The urban unemployed and their families in late Victorian Canada*, ed. P. A. Baskerville and E. W. Sager, 195–216. Toronto: University of Toronto Press.
- Block, W., and W. Thomas. 2003. Implementing the Data Documentation Initiative at the Minnesota Population Center. *Historical Methods* 36 (2): 97–101.
- Brookes, A. A. 1976. Doing the best I can: The taking of the 1861 New Brunswick census. *Histoire sociale/Social History* 9 (17): 70–91.
- Coats, R. H. 1946. Beginnings in Canadian statistics. *Canadian Historical Review* 37: 109–30.
- Curtis, B. 1994. On the local construction of statistical knowledge: Making up the 1861 census of the Canadas. *Journal of Historical Sociology* 7 (4): 416–34.
- . 2001. *The politics of population: State formation, statistics, and the census of Canada, 1840–1875*. Toronto: University of Toronto Press.
- Darroch, G., and L. Soltow. 1994. *Property and inequality in Victorian Ontario: Structural patterns and cultural communities in the 1871 census*. Toronto: University of Toronto Press.
- Dunae, P. A. 1998. Making the 1891 census in British Columbia. *Histoire sociale/Social History* 31 (62): 22–39.
- Fortier, N. 1984. Les recensements canadiens et l'étude de l'agriculture québécoise, 1852–1901. *Histoire sociale/Social History* 17 (33–34): 257–87.
- Gaffield, C. 1991. Children, schooling, and family reproduction in nineteenth-century Ontario. *Canadian Historical Review* 72: 157–91.
- . 2005. Evidence of what? Changing answers to the question of historical sources as illustrated by research using the census. In *Building new bridges: Sources, methods, and interdisciplinarity*, ed. J. Keshen and S. Perrier, 265–74. Ottawa: University of Ottawa Press.
- Gagan, D. 1981. *Hopeful travellers: Families, land, and social change*

- in mid-Victorian Peel County, Canada West*. Toronto: University of Toronto Press.
- Hacker, J. D., and C. A. Fitch, eds. 2003. Building historical data infrastructure: New projects of the Minnesota Population Center, Part 2. *Historical Methods* 36 (2): 60–101.
- Hacking, I. 1990. *The taming of chance*. Cambridge: Cambridge University Press.
- Hall, P. K., R. McCaa, and G. Thorvaldsen, eds. 2000. *Handbook of international historical microdata for population research: A project of IMAG, the international microdata access group*. Minneapolis: Minnesota Population Center.
- Igartua, J. E., ed. 1988. Les bases de données historiques: L'expérience canadienne depuis quinze ans [Historical databases: The Canadian experience after fifteen years]. *Histoire sociale/Social History* 21 (42): 283–387.
- Inwood, K., and R. Reid, eds. 1995. L'utilisation des données manuscrites de recensement dans la recherche historique [Introduction: The use of census manuscript data for historical research]. *Histoire sociale/Social History* 28 (56): 301–11.
- Katz, M. B. 1975. *The people of Hamilton, Canada West: Family and class in a mid-nineteenth-century city*. Cambridge: Harvard University Press.
- Kertzer, D. I., and D. Arel, eds. 2002. *Census and identity: The politics of race, ethnicity, and language in national censuses*. Cambridge: Cambridge University Press.
- Nilsen, K. 2001. *The impact of information policy: Measuring the effects of the commercialization of Canadian government statistics*. Westport, CT: Ablex.
- Olssen, E., and M. Hickey. 2005. *Class and occupation: The New Zealand reality*. Dunedin, New Zealand: University of Otago Press.
- Porter, T. M. 1986. *The rise of statistical thinking, 1820–1900*. Princeton: Princeton University Press.
- Sager, E. W., and P. Baskerville, eds. 2000. The Canadian Families Project: Exploring the 1901 Census of Canada. *Historical Methods* 33 (4): 179–263.
- . 2007. *Household counts: Canadian households and families in 1901*. Toronto: University of Toronto Press.
- Swierenga, R. P. 1990. Historians and the census: The historiography of census research. *Annals of Iowa* 50:650–73.
- Trudel, M. 1973. *La population du Canada en 1663*. Montreal: Fides.
- Van Loon, R., L. Marsden, C. Gaffield, J. McCamus, and G. La Forest. 2000. *Report of the expert panel on access to historical census records*. Ottawa: Department of Industry.
- Winchester, I. 1970. The linkage of historical records by man and computer: Techniques and problems. *Journal of Interdisciplinary History* 1:107–24.

Subscribe Today and Access Content Online!

Historical Methods reaches an international audience of historians and other social scientists concerned with historical problems. Features include the following: “Evidence Matters” emphasizes how to find, decipher, and analyze evidence whether or not that evidence is meant to be quantified. “Database Developments” announces major new public databases or large alterations in older ones, discusses innovative ways to organize them, and explains new ways of categorizing information. “Perfecting Data” addresses generic deficiencies in historical data and suggests ways to alleviate them. “Scholarly Incursions” includes bold cross-disciplinary approaches intended to shake up two or more fields of study. *Historical Methods* has also initiated an annual issue devoted entirely to lengthy reviews of recent books, emphasizing their methodological and theoretical aspects.



Regular Annual Subscription Rates:
Individual: \$68 online only, \$71 print and online
Institutional: \$173 online only, \$173 print only, \$208 print and online
 Add \$14 for postage outside the U.S.



1319 Eighteenth Street, NW
 Washington, DC 20036-1802
 P. 800.365.9753 ■ F. 202.293.6130
 P. 202.296.6267 ■ www.heldref.org

**CALL OR VISIT US
 ONLINE TO SUBSCRIBE!**

Libraries may order through subscription agents.