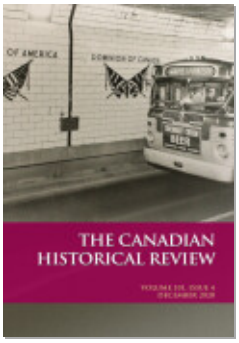
The Return of Quantitative Approaches to Canadian History

Peter Baskerville, Kris Inwood



The Canadian Historical Review, Volume 101, Issue 4, December 2020, pp. 585-601 (Article)

Published by University of Toronto Press

PETER BASKERVILLE and KRIS INWOOD

The Return of Quantitative Approaches to Canadian History

Abstract: *An important kind of digital-based research – the use of quantitative sources and analysis – expanded in Canadian history journals during the 1970s and 1980s, declined in the 1990s, and then has increased again since the early years of the new millennium. A wide variety of sources are being digitized and used for historical research that is wholly or, more commonly, partially quantitative in nature. The Canadian census is the most widely used of these sources. Access to census resources will improve considerably with the completion of a project currently in progress, which is entitled* The Canadian Peoples/Les populations canadiennes, *https://thecanadianpeoples.com.*

Keywords: digital, quantitative, census, historical methods, Canadian historiography

Résumé : *Un important type de recherche numérique – l'utilisation de sources quantitatives et leur analyse – a pris de l'ampleur dans les revues d'histoire canadiennes au cours des années 1970 et 1980 et a connu un nouvel essor depuis le début du nouveau millénaire, après avoir enregistré un recul dans les années 1990. On numérise actuellement toute une gamme de sources qui servent à la recherche historique, qu'elle soit entièrement ou, le plus souvent, partiellement quantitative. Les recensements canadiens sont la plus utilisée d'entre elles. L'accès aux richesses des recensements s'améliorera considérablement au terme d'un projet en cours intitulé* The Canadian Peoples/Les populations canadiennes, *https://thecanadianpeoples.com.*

Mots clés : numérique, quantitatif, recensement, méthodes historiques, historiographie canadienne

INTRODUCTION

Digital data have arrived. "The next big idea in language, history and the arts," according to Patricia Cohen, is "data."[1] If Cohen is even partially correct, a

1 Patricia Cohen, "Digital Keys for Unlocking the Humanities' Riches," *New York Times*, 16 November 2010.

 doi: 10.3138/chr-2020-0022

central question for historians today is how best to develop, understand, and make use of abundant digital data. Of course, there can be no single answer because digital information takes many forms. Words, sound, pictures, maps, and numbers are all increasingly accessible in digital form; each is used with its own constellation of tools and methods. Some of this information was "born digital," while other sources are digitized decades or even centuries after they were created. As historians, how can we best harness this diversity of new digital information in our quest to understand historical change? We argue that within the discipline of history, and the humanities more generally, the challenge of data abundance invites a paradigm shift in the way research is conducted and in our research tools. The methods and tools needed to make sense of digital media can take several forms, but we stress what we think is often overlooked – namely, the centrality of counting and the construction of structured databases.

At some level, all humanists, of course, count.[2] But, for many, this process is implicit rather than explicit. Taking full advantage of the new sources requires a modicum of familiarity with database construction, management, and analysis. Also helpful is some ability to code or program in languages originating with computing science or, at a minimum, familiarity with specialized software.[3] Many of these data-related skills also improve employment prospects outside of the academy. Careers after graduation are increasingly facilitated by expertise in the management of large masses of complex data and the capacity to interpret them critically, place them in context, identify deeper significance, and communicate results to a general audience. Admittedly, one complication of the new digital history is that the lone scholar increasingly is supplanted by the "collaboratory," a laboratory without walls that facilitates collective and cooperative approaches to research.[4] The diversity of skills and resources available to interdisciplinary teams or networks better equips them to meet the challenges of data abundance. We illustrate these issues with an impending innovation in access to Canada's historical census information after first reviewing the state of routinely generated historical digital data in Canadian history.

2 Scholars count all of the reasons that quantification has long been useful. See Alfred W. Crosby, *The Measure of Reality: Quantification in Western Europe, 1250–1600* (Cambridge, UK: Cambridge University Press, 1988); Patricia Cohen, *A Calculating People: The Spread of Numeracy in Early America* (Chicago: University of Chicago Press, 1982); Patricia Cohen, "The Emergence of Numeracy," in *Mathematics and Democracy: The Case for Quantitative Literacy*, ed. Lynn Arthur Steen (New York: Woodrow Wilson Foundation, 2001), 23–30.

3 We are thinking here of databases, statistical research, geographical information systems, image editing, text-editing software, and so on. Even spreadsheet software, which is ubiquitous and easy to use, is highly effective for many historical research purposes.

4 Stefan Dormans and Jan Kok, "An Alternative Approach to Large Historical Databases: Exploring Best Practices with Collaboratories," *Historical Methods* 43, no. 3 (2010): 97–107.

HISTORIOGRAPHICAL ANALYSIS OF PUBLISHING TRENDS

The use of quantitative methods by Canadian historians is to a great degree a story of two solitudes. In Quebec, the understanding of history draws on a strong tradition of demographic analysis. This tradition was rooted in nationalist preoccupations with the peopling of New France and was facilitated by a rich body of population records, including parish-level birth, marriage, and death records. During the third quarter of the twentieth century, Quebec demographers learned the techniques of family reconstitution from their contemporaries in France and began massive projects to reconstruct the entire populations of New France and the Saguenay region.[5] Particularly influential in the Quebec demographic tradition was the work of distinguished researchers at the Université de Montréal, l'Université du Québec à Chicoutimi, and the Centre interuniversitaire d'études québécoises.[6] Demographic analysis revealed the adaptive capacities of individuals and groups in the new world. Structural and institutional constraints persisted into the twentieth century, as did the adaptive capacities of the population.

In English Canada, demographic analysis of large populations has never been so intensive, in spite of the challenges afforded by fertility decline in Upper Canada/Ontario and massive net emigration later in the nineteenth century. Structural analysis during the 1970s was largely the work of social historians in collaborative projects that made extensive use of census data.[7] The Hamilton Social History Project, led by Michael Katz, produced the first major studies of urban families in English Canada.[8] Michael Ornstein and Gordon Darroch constructed a sample from the 1871 Canadian census and then a linked 1861–71 sample to investigate occupational mobility, ethnic stratification, and social inequality.[9] David Gagan

5 Jacques Henripin, *La Population canadienne au début du 18e siècle: nuptialité, fécondité, mortalité infantile* (Paris: Presses universitaires de France, 1954); Hubert Charbonneau, *La population du Québec: études rétrospectives* (Montreal: Les éditions du Boréal express, 1973); Jacques Légaré, "A Population Register for Canada under the French Regime: Context, Scope, Content and Applications," *Canadian Studies in Population* 15, no. 1 (1988): 1–16; Gérard Bouchard, "Family Reproduction in New Rural Areas: Outline of a North American Model," *Canadian Historical Review* 75, no. 4 (1994): 475–510; Gérard Bouchard, *Quelques Arpents d'Amérique: Population, Économie, Famille au Saguenay, 1838–1971* (Montreal: Les Éditions de Boréal, 1996).

6 Programme de recherche en démographie historique, https://www.prdh-igd.com; Centre interuniversitaire d'études québécoises, https://www.cieq.ca; BALSAC, http://balsac.uqac.ca; http://balsac.uqac.ca/fichier-balsac/historique/.

7 Analysis and a broader contextualization of these developments are provided by Chad Gaffield, "Clio and Computers in Canada and Beyond: Contested Past, Promising Present, Uncertain Future," in this issue.

8 Michael Katz, *The People of Hamilton, Canada West: Family and Class in a Mid-Nineteenth Century Canadian City* (Cambridge, UK: Cambridge University Press, 1975); Michael Katz, J. Doucett, and M. Stern, *The Social Organization of Early Industrial Capitalism* (Cambridge, UK: Cambridge University Press, 1982).

9 Gordon Darroch and Michael Ornstein, "Ethnicity and Class: Transitions over a Decade: Ontario 1861–71," *Canadian Historical Association Historical Papers* 19 (1984): 111–37; Gordon Darroch and Michael Ornstein, "Family and Household in Nineteenth-Century Canada: Regional Patterns and Regional Economies," *Journal of Family History* 9, no. 2 (1984): 158–717.

produced the first major study of English Canadian rural families – work that is still being debated.[10] The important influences here came from the United States and the United Kingdom. The central questions were about social and geographic mobility, family structures, and life cycles. These projects were also among the first to integrate census data with other sources for the study of family strategies: wills, school attendance records, assessment rolls, city directories, letters, and diaries. Such projects emerged from the trans-Atlantic fixation with urbanization, industrialization, and class formation in the nineteenth century.

The influence of seminal research publications in the 1970s (and earlier for Quebec) contributed to the appearance of new journals such as *Histoire sociale/Social History*, *Labour/Le Travail*, the reborn *Acadiensis*, and an increasing appreciation for the contributions of quantitative evidence. By the mid-1980s, the quantitative turn in Canadian history was well advanced, as is evident from a review of scholarly journals partly or entirely dedicated to Canadian history. The content analysis summarized below suggests three distinctive epochs in the writing of Canadian history over the last five decades. We assess the contribution of systematically generated or quantitative sources through an identification of journal articles that employ data in a visible and significant manner. The journals we examine will be familiar to all Canadian historians: *The Canadian Historical Review*, *Acadiensis*, *Revue de l'histoire de l'Amérique française*, *Ontario History*, BC *Studies*, *Journal of the Canadian Historical Association* (formerly *Historical Papers*), *Histoire sociale/Social History*, and *Labour/Le Travail*. We exclude articles that are not directly focused on Canada, which are not primarily historical, and editors' notes, bibliographies, review articles (unless they are particularly substantial), comments, and replies (unless the reply presents information clearly absent from the original article). Overall, we examined 4,384 articles over the fifty years from 1969 to 2018, which comprised about fifty articles per year at the beginning of this period and rose to more than one hundred articles per year at its end (Figure 1). Most of the growth in the number of publications was experienced during the 1970s.

A few caveats are in order. The presence of a single table or a few text references does not suffice to be designated quantitative; rather, we look for a significant presence of numbers in the text, tables, graphs, or maps. Admittedly, this is complicated insofar as the argument of a paper may rest almost entirely on quantitative sources, even though few, if any, numbers are visible.[11] Other

10 David Gagan, *Hopeful Travellers: Families, Land and Social Change in Mid-Victorian Peel County, Canada West* (Toronto: University of Toronto Press, 1981). For examples of such discussion, see George Emery and José Igartua, "David Gagan's 'The Critical Years in Rural Canada West': A Critique of Methodology and Model," *Canadian Historical Review* 62, no. 2 (1981): 186–96; David Gagan, "Under the Lamp Post: A Reply to Emery and Igartua," *Canadian Historical Review* 62, no. 2 (1981): 197–206; Peter Baskerville, "Mortgaging the Rural Homestead: Common Perceptions Reconsidered, the Case of Logan Township, Perth County, Ontario, 1850–1911" (Paper presented at the Social Science History Association Annual Meeting, Vancouver, British Columbia, 2012); Peter A. Russell, *How Agriculture Made Canada: Farming in the Nineteenth Century* (Montreal and Kingston: McGill-Queen's University Press, 2012), 142–68.

11 Underlying this ambiguity lies a fundamental problem of distinguishing the implicit from the explicit use of numbers. Peter Laslett, *Family Life and Illicit Love*

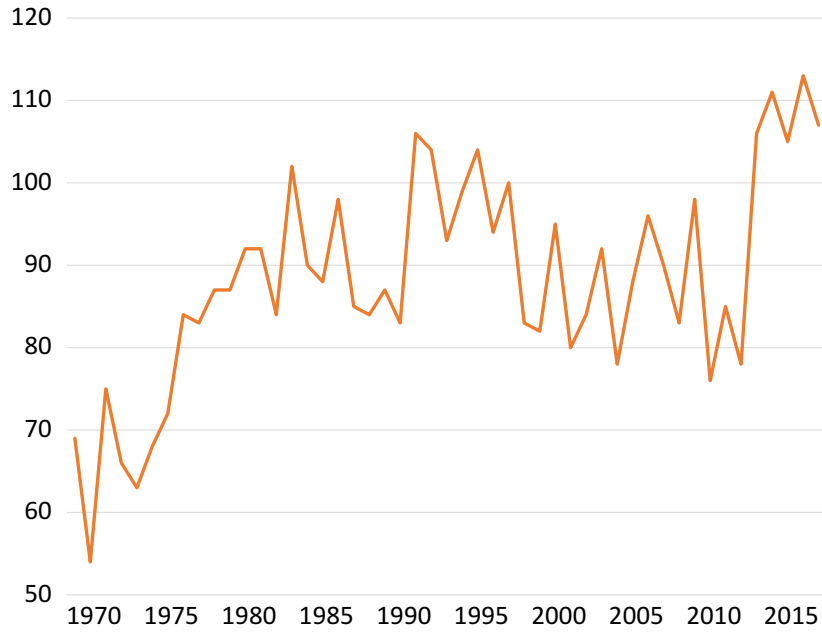


FIGURE 1 Number of Canadian history articles published annually in eight journals. *Source: The Canadian Historical Review, Acadiensis, Revue de l'histoire de l'Amérique française, Ontario History,* BC *Studies, Journal of the Canadian Historical Association* (formerly *Historical Papers*), *Histoire sociale/Social History*, and *Labour/Le Travail.*

papers, thankfully limited in number, contain graphs and tables with little connection to the central point of the article. Of course, research that employs quantitative evidence is seldom purely quantitative. Rather, as with any single source, a wider range of information and careful qualitative reasoning almost always is needed to develop a credible interpretation and extract full information from the source.[12] There is no easy or simple way to summarize the demarcation except to say that we have examined each article individually using our judgement as Canadian historians to assess the importance of the numbers. Similarly, we use our judgement to distinguish the importance (or not) of Canadian history in contributions to international, global, or comparative research. It must be admitted that a review of these eight journals misses Canadian work in international journals and ignores historical or long-term analyses appearing in journals of other disciplines, monographs, and collections of scholarly articles.[13] Neither have we

in Earlier Generations: Essays in Historical Sociology (Cambridge, UK: Cambridge University Press, 1977), 7.

12 Among the many, many books and articles that may be cited here, it may be useful to mention a particularly effective combination of qualitative and quantitative reasoning and sources. Chad Gaffield, *Language, Schooling and Cultural Conflict: The Origins of the French-Language Controversy in Ontario* (Montreal and Kingston: McGill-Queen's University Press, 1987).

13 See Table 1 for an indication of such activity following 2008.

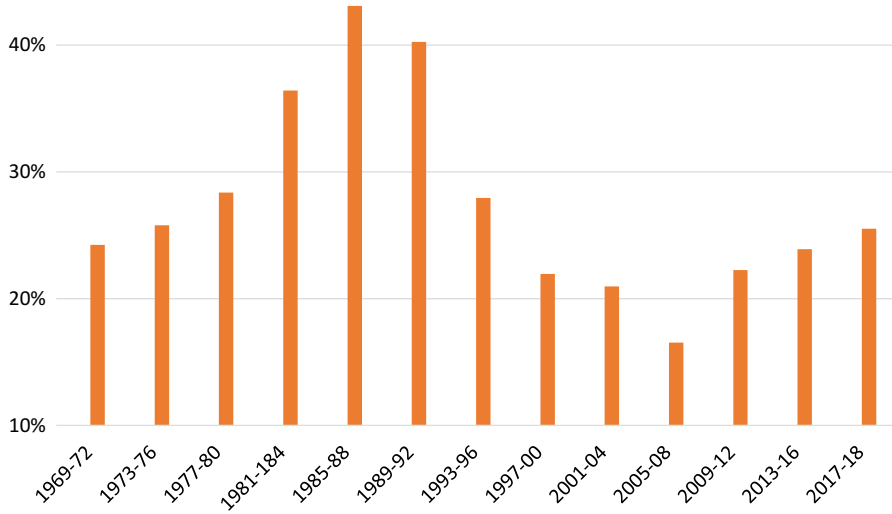


FIGURE 2 Share of articles with quantitative evidence in eight Canadian history journals. *Source: The Canadian Historical Review, Acadiensis, Revue de l'histoire de l'Amérique française, Ontario History,* BC *Studies, Journal of the Canadian Historical Association* (formerly *Historical Papers*), *Histoire sociale/Social History,* and *Labour/Le Travail.*

been able to consider the increasingly important communication of historical debate and knowledge through web pages, blogs, and social media.

The somewhat stylized overview in Figure 2 suggests a dramatic expansion of the role of counting in Canadian history journals during the 1970s and 1980s.[14] Roughly one-quarter of all of our articles visibly relied on quantitative evidence in the early 1970s. By the late 1980s, the share had risen to nearly one-half.[15] The simultaneous growth of Canadian history, as suggested by Figure 1, implies an even greater expansion in the overall volume of research using quantitative evidence. Not coincidentally, of course, precisely in this period most universities began to offer centralized computing capacity to its researchers, and historians quickly adopted personal computers once they became affordable.[16] By the late 1980s, entire issues of *Histoire sociale/Social History* and *Labour/Le Travail* and even the Canadian Historical Association's *Historical Papers* were dominated by articles using quantitative evidence. Important roundtable discussions remarked and reflected on the development.[17] A number of the winners and honourable mentions

14 Our identification of quantitative is broader than that of Peter George and Ernest Oksanen, "Recent Developments in the Quantification of Canadian Economic History," *Histoire sociale/Social History* 4 (1969): 76–98, and similar to that of Kris Inwood, "The Promise and Problems of Quantitative Evidence in Canadian History," *Histoire sociale/Social History* 27, no. 53 (1994): 139–46.

15 Single-year averages of forty-nine percent in 1987 and 1989 mark the peak.

16 The importance of an evolution in accessible computing power is highlighted by Gaffield, "Clio and Computers."

17 José Igartua et al, "Historical Databases: The Canadian Experience," *Histoire sociale/Social History* 21, no. 42 (1988): 283–318.

for the Best Scholarly Book in Canadian History Prize (formerly the Macdonald Prize) during the 1980s and early 1990s deployed quantitative evidence in a significant manner.[18] The use of numeric data in those years seemed to be at the cutting edge of historical development throughout Western Europe and the United States.[19] Not surprisingly, it also moved into the mainstream of Canadian history.

And, then, the tide receded. Canadian historians, like their colleagues elsewhere, increasingly turned to phenomena that were not easily counted and to analytical subtleties for which simple categorizations and "broad, but thin" quantitative evidence were not helpful.[20] Between 1990 and 2005, the share of articles using quantitative evidence declined continuously under the influence of the "cultural turn," and a growing appreciation for postmodern epistemologies in the humanities generally increased.[21] The decline continued until 2004–5, when the share of quantitative articles fell below fifteen percent.[22] The saga of shifting conceptual frameworks and methodologies is well known. Perhaps less recognized is the rebound since 2005 in the willingness of Canadian historians to entertain and report their use of quantitative evidence. Over the ten years since the nadir in 2005, the quantitative-using share of research articles rose steadily. It has now returned to its level of circa 1970, which is about twenty-five percent. To a large extent, this expansion parallels a change that is visible in journals published in the United States.[23] The most frequently used source of quantitative information for Canadian history, both before and after 2005, has been the census.

We look more closely at the experience of these ten years in Table 1. From twenty to thirty percent of the articles published in Canadian journals make use of quantitative evidence – apart from *The Canadian Historical Review* and *Ontario History*, which continue to be the least quantitative. Differences among the other six journals are now smaller than they used to be. Canadian history also appears in international journals. A final line in Table 1 reports on ten international journals that together account for several Canadian history papers with

18 Authors and titles are listed at Canadian Historical Association, "CHA Prizes," https://cha-shc.ca/english/what-we-do/prizes/the-cha-best-scholarly-book-in-canadian-history-prize-formerly-known-as-the-sir-john-a-macdonald-prize-from-1977-to-2018.htm.

19 Steven Ruggles, "The Revival of Quantification," *Social Science History* (forthcoming); Ruggles, "Disciplinary Trends in Quantitative Social Science History" (Paper presented at Social Science History Association Annual Meeting, Chicago, 2004).

20 Among the many accounts of the "cultural turn" and discourse analysis in Canadian history, see the reflections from different perspectives by Mariana Valverde, "Some Remarks on the Rise and Fall of Discourse Analysis," *Histoire sociale/Social History* 33 (2000): 59–77; Joan Sangster, "Manufacturing Consent in Peterborough" in *Through Feminist Eyes: Essays on Canadian Women's History* (Edmonton: Athabasca University Press, 2011), 127–34.

21 John F. Reynolds, "Do Historians Count Anymore?: The Status of Quantitative Methods in History, 1975–1995," *Historical Methods* 31, no. 4 (1988): 141–8.

22 The decline was experienced by all of our journals, although fluctuation was more modest for the *Canadian Historical Review* and *Revue de l'histoire de l'Amérique français*.

23 Ruggles, "Revival of Quantification."

TABLE 1
Canadian articles using quantitative evidence in select journals, 2009–18

Source	Articles with quantitative evidence	Percentage of journal's Canadian articles
Canadian Historical Review	19	13
Ontario History	15	17
Revue d'histoire de l'Amérique française	20	20
Acadiensis	28	25
Journal of the Canadian Historical Association	24	26
Histoire sociale/Social History	38	26
BC Studies	47	28
Labour/Le Travail	37	32
Ten international journals*	39	66

Note: * *Historical Methods*, *History of the Family*, *Social Science History*, *Journal of Family History*, *Journal of Interdisciplinary History*, *Journal of Social History*, *Journal of Historical Sociology*, *Journal of Historical Geography*, *Journal of Economic History*, and *Annales de démographie historique*.

quantitative evidence. The higher proportion of quantitative material may not be typical of all international journals, although quantification and the systematic use of structured numeric databases clearly resonates with an international community of practitioners in a number of important fields. It is at this level of innovation that Canadian historians are at the cutting edge.

What exactly do we mean by "level of innovation?" Activity at this knowledge frontier reflects the paradigm shift that we argue must take place within the discipline of history and the humanities more generally. Scholars from history and the allied disciplines are harnessing twenty-first-century computing power, multidisciplinary expertise, and sophisticated approaches to constructing, managing, storing, and analyzing large amounts of information. As in the sciences, scholars in the humanities are also engaged in unprecedented efforts at knowledge mobilization. Research questions and agendas interact with technology and methods, and these interactions change the practice of history and our intellectual horizons. The movement to make research data more accessible is gaining momentum.[24] Open domain databases can be seen as creative engines that, by restructuring traditional archival boundaries, enable new ways of thinking and new approaches to research.

Our review of Canadian history journals over the last fifty years suggests, notwithstanding the ebb and flow of intellectual fashion, that Canadian history will continue to be informed by quantitative evidence and by reasoning based on things we can count. For most topics, historians will continue to use a range

24 "Tri-Agency Statement of Principles on Digital Data Management," *Government of Canada*, http://www.science.gc.ca/eic/site/063.nsf/eng/h_83F7624E.html?OpenDocument.

of sources, among which some may be quantitative, while others are not, as several recent articles in this journal illustrate.[25] The exceptions here are specialist articles in economic or demographic history[26] and spatial analysis using the fast-growing methodology of historical geographic information systems.[27] To a large extent, Canadian historians have converged on a position initially received as a minority view when it was articulated twenty years ago by Eric Sager:

> Where the dynamics of change are the target, and where the historian is seeking the conditions of change affecting large numbers of cases, quantitative measures and statements are almost inescapable. The careful historian will want to carry the appropriate tools and to be fully conscious of their uses and limits ... the exceptional or oppositional case cannot be discerned until a norm or trend is specified. ... Anecdotes add little in the way of substantive evidence, since counter-examples can almost always be found. ... Statistical methods and representations are a pre-eminent means by which the human sciences present human voices and actions.[28]

Broad acceptance of this perspective has prepared Canadian historians for the new digital sources now becoming available. Earlier generations fashioned

25 Jan Raska, "Humanitarian Gesture: Canada and the Tibetan Resettlement Program, 1971–5," *Canadian Historical Review* 97, no. 4 (2016): 546–75; Andrew Watson, "Pioneering a Rural Identity on the Canadian Shield: Tourism, Household Economies, and Poor Soils in Muskoka, Ontario, 1870–1900," *Canadian Historical Review* 98, no. 2 (2017): 261–93; Peter L. Twohig, "The Second 'Great Transformation': Renegotiating Nursing Practice in Ontario, 1945–70," *Canadian Historical Review* 99, no. 2 (2018): 169–95; Edward Dunsworth, "Race, Exclusion, and Archival Silences in the Seasonal Migration of Tobacco Workers from the Southern United States to Ontario," *Canadian Historical Review* 99, no. 4 (2018): 563–93.

26 Sylvie Taschereau and Yvan Rousseau, "The Hidden Face of Consumption: Extending Credit to the Urban Masses in Montreal (1920s to 40s)," *Canadian Historical Review* 100, no. 4 (2019): 509–39. Even in specialist fields, the publishing of books typically relies on a broader range of sources. For example, Peter Baskerville, *A Silent Revolution?: Gender and Wealth in English Canada, 1860–1930* (Kingston and Montreal: McGill-Queens University Press, 2008); John Belshaw, *Becoming British Columbia: A Population History* (Vancouver: UBC Press, 2009).

27 Jennifer Bonnell and Marcel Fortin, eds., *Historical GIS Research in* Canada (Calgary: University of Calgary Press, 2014); Patrick Dunae et al., "Dwelling Places and Social Spaces: Revealing the Environment of Urban Workers in Victoria Using Historical GIS," *Labour/Le Travail* 72 (2013): 37–73; Sasha Mullaly and Siobhan Hanratty, "Visualizing the Past: Mapping, GIS, and Teaching Historical Consciousness," *Active History*, 13 July 2016; "Montreal: L'avenir du passe," http://www.mun.ca/mapm. For a comprehensive list (approximately fifty-five pages and growing) of historical geographic information system studies covering an international literature that lists at least fifty-five contributions on Canada, see especially Historical Geographic Information System Lab, University of Saskatchewan, *Bibliography*, https://hgis.usask.ca/resources/bibliography.php#_Historical_GIS_A.

28 Eric Sager, "Employment Contracts in Merchant Shipping: An Argument for Social Science History," in *On the Case: Explorations in Social History*, ed. Franca Iacovetta and Wendy Mitchinson (Toronto: University of Toronto Press, 1998), 59–61.

databases from individual-level records chronicling birth, marriage, and death, factory workers, rural crafts, maritime employment, strikes, property ownership and tenancy, gender, investments, ethnicity, health, religion, and so on in order to explore relationships and construct collective descriptions.[29] More recently, historians have developed databases to investigate soldiers,[30] prisoners,[31] homesteaders,[32] rural consumption and shopping,[33] commodity trade,[34] and no doubt many more. Another growing genre is the qualitative appraisal of quantitative sources. It is now widely accepted that apparently systematic sources, like any

29 Charbonneau, *La population du Québec*; Bouchard, *Quelques Arpents d'Amérique*; Gregory Kealey, *Toronto Workers Respond to Industrial Capitalism, 1867–1892* (Toronto: University of Toronto Press, 1980); Kris Inwood and Phyllis Wagg, "The Survival of Handloom Weaving in Rural Canada circa 1870," *Journal of Economic History* 54 (1993): 346–58; Eric Sager, *Seafaring Labour: The Merchant Marine of Atlantic Canada, 1820–1914* (Montreal and Kingston: McGill-Queen's University Press, 1989); Peter Oliver, *Terror to Evil-Doers: Prisons and Punishments in Nineteenth-Century Ontario* (Toronto: University of Toronto Press, 1998); Douglas Cruikshank and Gregory S. Kealey, "Strikes in Canada, 1891–1950: I. Analysis," *Labour/Le Travail* 20 (1987): 85–145; Gordon Darroch and Lee Soltow, *Property and Inequality in Victorian Ontario: Structural Patterns and Cultural Communities in the 1871 Census* (Toronto: University of Toronto Press, 1984); Catharine Wilson, *Tenants in Time: Family Strategies, Land, and Liberalism in Upper Canada 1799–1871* (Kingston and Montreal: McGill-Queen's University Press, 2009); Peter Baskerville, "Women and Investment in Late-Nineteenth-Century Urban Canada: Victoria and Hamilton, 1880–1901," *Canadian Historical Review* 80, no. 2 (1999): 191–218; Ken Sylvester, "All Things Being Equal: Land Ownership and Ethnicity in Rural Canada, 1901," *Histoire sociale/Social History* 34 (2001): 35–59; John Cranfield and Kris Inwood, "The Great Transformation: A Long-Run Perspective on Physical Well-Being in Canada," *Economics and Human Biology* 5 (2007): 204–28; Peter Baskerville, "Did Religion Matter? Religion and Wealth in Urban Canada at the Turn of the Twentieth Century: An Exploratory Study," *Histoire sociale/Social History* 34 (2001): 61–95.

30 Nic Clarke, *Unwanted Warriors: Rejected Volunteers of the Canadian Expeditionary Force* (Vancouver: UBC Press, 2016); Kris Inwood and J. Andrew Ross, "Big Data and the Military: First World War Personnel Records in Australia, Britain, Canada, New Zealand and British Africa," *Australian Historical Studies* 47 (2016): 430–42.

31 Donald Fyson and François Fenchel, "Prison Registers, Their Possibilities and Their Pitfalls: The Case of Local Prisons in Nineteenth Century Quebec," *History of the Family* 20, no. 2 (2015): 163–88; Kris Inwood, "'Indians Are the Majority of the Prisoners'? Historical Variations in Incarceration Rates for Indigenous Women and Men in British Columbia," *Howard Journal of Criminal Justice* (forthcoming).

32 Luiza Antonie et al., "Population Analysis of the Settlement Movement in Western Canada," *International Journal of Population Data Science* 3 (2018): 275–275; Peter Baskerville, "The Last Best West: Alberta's Homesteaders on the Move, 1871–1916, an Exploratory Study" (Paper presented to the Canadian Historical Association Annual Meeting, Ottawa, June 2015).

33 Douglas McCalla, *Consumers in the Bush: Shopping in Rural Upper Canada* (Kingston and Montreal: McGill-Queen's University Press, 2015).

34 Jim Clifford et al., "Geoparsing History: Locating Commodities in Ten Million Pages of Nineteenth Century Sources" *Historical Methods* 49, no. 3 (2016): 115–31; Uta Hinrichs et al., "Trading Consequences: A Case Study of Combining Text Mining and Visualization to Facilitate Document Exploration," *Digital Scholarship in the Humanities* 30, no. S1 (2015): i50–i75.

other, must be assessed carefully in order to uncover the social constructions and at times idiosyncratic circumstances that limit and condition their uses.[35]

The most exciting development of the last twenty years, however, has been the construction of representative samples from Canada's historical censuses, which are widely recognized for their outstanding quality, comprehensive coverage, and unique features of broad international interest. The evolution and use of census resources is a particular example that nonetheless reflects a broader transformation being experienced in multiple genres with diverse sources.

The Contribution of the Canadian Census

The census is an unparalleled historical source for population and geographic coverage. The 1871 Canadian enumeration was the world's most carefully undertaken national census to that date. The 1901 Canadian census is the most careful enumeration in the Americas of Indigenous and mixed-ancestry populations. Our 1901, 1911, and 1921 enumerations were the first in the world to report employment earnings. All Canadian censuses recorded religion and ethnicity. Several reported skin colour (1848, 1851, 1861, and 1901). Mother tongue is reported beginning in 1901. No other country inventoried these characteristics as early and as systematically as did Canada. The census is the only Canadian source before the mid-twentieth century that is not highly selective in representing populations in terms of class, ethnicity, gender, education, or location. These microdata on individual lives (some censuses provide answers for over forty questions for each individual) are, indeed, the only surviving record left by most Canadians in earlier decades. To construct, manage, and analyze such large aggregations of data required collaborative organization, which brought together scholars from history, geography, sociology, demography, and economics.

A manuscript sample of the 1871 Canadian census prepared by Gordon Darroch and Michael Ornstein was unveiled in this journal in 1980.[36] Since then,

35 Lisa Dillon, Brian Gratton, and Jon Moen, "Retirement at the Turn of the Twentieth Century: A Canadian Perspective," *Canadian Historical Review* 91, no. 1 (2010): 27–59; Chad Gaffield, Byron Moldofsky, and Katharine Rollwagen, "Do Not Use for Comparison with Other Censuses: Identity, Politics and Languages Commonly Spoken in 1911 Canada," in *The Dawn of Canada's Century: Hidden Histories*, ed. Gordon Darroch (Montreal and Kingston: McGill-Queen's University Press, 2014), 92–123; Michelle Hamilton, "Anyone Not on the List Might as Well Be Dead: Aboriginal Peoples and the Censuses of Canada, 1851–1916," *Journal of Canadian Historical Association* 18 (2007): 57–79; Michelle Hamilton and Kris Inwood, "The Aboriginal Population and the 1891 Census of Canada," in *Indigenous Peoples and Demography: The Complex Relation between Identity and Statistics*, ed. Per Axelsson and Peter Skold (New York: Bergahn, 2011), 95–116; Kris Inwood and Richard Reid, "Gender and Occupational Identity in a Canadian Census," *Historical Methods* 32 (2001): 57–70; Patrick Dunae, "Sex, Charades and Census Records: Locating Female Sex Workers in a Victorian City," *Histoire sociale/Social History* 42 (2009): 267–97.

36 Gordon Darroch and Michael Ornstein, "Ethnicity and Occupational Structure in Canada in 1871: The Vertical Mosaic in Historical Perspective," *Canadian Historical Review* 61, no. 3 (1980): 305–33.

historians have digitized samples of all Canadian census enumerations beginning with 1852. The proportion of records in each enumeration selected for digitization varies between twenty percent for 1852 and three percent for 1931, 1941, and 1951.[37] Most of these databases were produced by collaboratories involving many individuals to ensure interoperability across national census years. Only the 1871 census sample existed until 2003, in which year a multi-institutional team headquartered at the University of Victoria released a sample of the 1901 census. In 2009, another team headed by the University of Ottawa sampled the 1911, 1921, 1931, 1941, and 1951 enumerations.[38] Projects at the University of Guelph and Université de Montréal subsequently constructed samples from the Canadian census enumerations of 1852, 1861, and 1891. Separately, the University of Ottawa organized the construction of a database containing all records from the 1881 enumeration.[39] The most recent addition is a digitization of the entire 1831 census of Lower Canada at the Université de Montréal.[40] These databases, made available during the last fifteen years, can be used together with Statistics Canada's anonymized census samples from 1971 and subsequent years to support detailed investigations of Canadian population and society at regular intervals over the last 165 years.[41]

As Chad Gaffield points out, careful analysis of these data enables interpretations in which large-scale social, economic, and cultural change emerges from multiple individual decisions and actions compounded across the entire population. Census data illuminate not only the collective experience of social

37 Peter Baskerville and Eric Sager, eds., *Household Counts* (Toronto: University of Toronto Press, 2007); Gordon Darroch, ed., *The Dawn of Canada's Century* (Kingston and Montreal: McGill-Queens University Press, 2014); Kris Inwood and Kevin James, "Une ressource numérique pour l'analyse historique: le recensement canadien de 1891," *Cahiers québécois de démographie* 34 (2005): 315–29; Lisa Dillon and Katrina Joubert, "Dans les pas des recenseurs: une analyse critique des dimensions géographiques et familiales du recensement canadien de 1852," *Cahiers québécois de démographie* 41 (2012): 299–339.

38 The 1911–51 databases are remarkable for their inclusion of quantitative contextual data, and geographic information system-based geo-referencing as well as the census data themselves. See Eric Sager and Peter Baskerville, "Canadian Historical Research and Pedagogy: A View from the Perspective of the Canadian Century Research Infrastructure Project," *Canadian Historical Review* 91, no. 3 (2010): 533–51; Chad Gaffield, "Conceptualizing and Constructing the Canadian Century Research Infrastructure," *Historical Methods* 40 (2007): 54–64; Marc St-Hilaire et al., "Geocoding and Mapping Historical Census Data: The Geographical Component of the Canadian Century Research Infrastructure," *Historical Methods* 40 (2007): 76–91; Claude Bellavance, France Normand, and Evelyn S. Ruppert, "Census in Context: Documenting and Understanding the Making of Early-Twentieth Century Canadian Censuses," *Historical Methods* 40 (2007): 92–103.

39 Lisa Dillon, "International Partners, Local Volunteers and Lots of Data: The 1881 Canadian Census Project," *History and Computing* 12 (2000): 163–76.

40 Isabelle Cherkesly, Lisa Dillon, and Alain Gagnon, "Creating the 1831 Canadian Census Database," *Historical Methods* 52 (2019): 110–27.

41 The 1871 and 1901 data were financed in large part by the Social Sciences and Humanities Research Council, while the Canadian Foundation for Innovation and its provincial counterparts funded the other census databases.

change but also the diversity of individual experience in particular times and places. Micro-level documentation of the lives of ordinary people in matters such as employment, marriage, fertility, and migration provides unprecedented insight into the broad patterns of historical transformation.[42] The paradigm of open collaborative interaction in the construction of these census databases has led, almost inevitably, to another level of innovation. These infrastructures have the capacity to transcend national boundaries and to interrogate the significance of those boundaries. Indeed, the creation of population databases is a global enterprise. Canadian census data have been constructed to fit easily into international data infrastructure headquartered at the University of Minnesota.[43] Thus, Canadian census data from 1852 to 2011 is already available on a comparable basis with data from ninety-three other countries. These are interoperable data on a truly global scale.[44]

The new historical census databases are beginning to have an impact on Canadian research. Admittedly there are barriers. The effective use of these digital data inevitably requires some degree of specialist knowledge about data manipulation and statistical inference in addition to the traditional historical skills of source criticism and evaluation. Perhaps because of this, the census samples have been most valuable for answering specialist questions about social structure, inequality, and economic and demographic behaviours that typically sit at the margins of historical understanding. Another great limitation is that most of the current historical census databases are merely samples. Specialist researchers can do a great deal with a five percent sample of an entire population, but it is of limited value for most historical purposes. To take one example, researchers might use the historical census samples to investigate the lives of Canadian Mennonites.[45] Unfortunately, the five percent sample available for 1901 misses most of the thirty-two thousand individuals identified as Mennonite in that year. While information about five percent of the Mennonites (or roughly sixteen hundred) in 1901 may be a useful resource, it is insufficient to identify patterns with any precision or to use the census in order to contrast this group with other religions. Neither is it possible to use the census in order to follow the lives of individual Mennonites from census to census on a large enough

42 Chad Gaffield, "Evidence of What? Changing Answers to the Question of Historical Sources as Illustrated by Research Using the Census," in *Building New Bridges – Batir de nouveaux ponts: Sources, Methods and Interdisciplinarity*, ed. Jeff Keshen and Sylvie Perrier (Ottawa: University of Ottawa Press, 2005), 265–74.

43 IPUMS International, "Harmonized International Census Data for Social Science and Health Research," https://international.ipums.org/international/.

44 Recent works exploiting globally integrated data include Lisa Dillon, *The Shady Side of Fifty: Age and Old Age in Late Victorian Canada and the United States* (Kingston and Montreal: McGill-Queen's University Press, 2008); Jason Long and Joseph Ferrie, "Intergenerational Occupational Mobility in Great Britain and the United States since 1850," *American Economic Review* 103 (2013): 1109–37.

45 Royden Loewen, *Family, Church and Market: Mennonite Communities in the Old and the New Worlds, 1850–1930* (Toronto: University of Toronto Press, 1993); Royden Loewen, *Hidden Worlds: Revisiting the Mennonite Migrants of the 1870s* (Winnipeg: University of Manitoba Press, 2001).

scale to construct a typical life pattern because the great majority of the individuals identified in the 1901 sample will not be found in the randomly selected four percent sample of the 1911 census or in subsequent samples. These, and many other standard historical research questions cannot be addressed with samples, even if they are randomly selected and perfectly representative of the broader population.

A new digital resource currently in preparation will remedy this deficiency. An interdisciplinary team of researchers at nine Canadian universities, in collaboration with the genealogical company Ancestry, will digitize, code, and make useable for research the personal, family, and household characteristics of every individual enumerated in each census from immediately before Confederation to shortly after the First World War.[46] The individual universities, the Canadian Foundation for Innovation, provincial research funds, and Ancestry provide the financing. This collaboration among private business, government, and academics illustrates the way in which research collaboratories in the world of big data cross hitherto seemingly impermeable boundaries. The Canadian Peoples/Les populations canadiennes (TCP/Lpc) will make available for research use about forty million records from every census between 1852 and 1926.[47] Returning to our example, after the new databases become accessible to researchers (in 2021 or 2022 if there are delays), it will be possible with minimal effort to analyze every enumerated characteristic for everyone identified as Mennonite in all of the census enumerations not restricted by privacy legislation. TCP/Lpc data will provide detailed information about, literally, everyone enumerated in Canada during its first sixty years. This will facilitate greatly two novel lines of research: the systematic examination of small minorities without losing sight of the aggregate experience and the study of large numbers of people through time in true longitudinal fashion.[48]

Following large numbers of people from one census to another (and to other sources) is now possible thanks to custom software that tracks individuals across successive decadal census intervals and to vital records (where they are available).[49] The software is one variant of large-scale text analysis that

46 The University of Guelph leads the team that includes the University of Alberta, the University of Saskatchewan, the University of Toronto, Queens University, the University of Ottawa, Université de Montréal, Université Laval, and the University of New Brunswick. Ancestry, https://www.ancestry.com.

47 The Canadian Peoples/Les populations canadiennes, https://thecanadianpeoples.com.

48 George C. Alter et al., "Longitudinal Analysis of Historical-Demographic Data," *Journal of Interdisciplinary History* 42 (2012): 503–17; Peter Baskerville and Kris Inwood, eds., *Lives in Transition: Longitudinal Analysis from Historical Sources* (Kingston and Montreal: McGill-Queen's University Press, 2015).

49 Luiza Antonie et al., "Tracking People over Time in 19th Century Canada," *Machine Learning* 96 (2014): 129–46; Luiza Antonie, Kris Inwood and J. Andrew Ross, "Dancing with Dirty Data: Problems in the Extraction of Life-Course Evidence from Historical Censuses," in *Population Reconstruction*, ed. G. Bloothooft et al. (New York: Springer, 2015), 217–41; Catalina Torres and Lisa Dillon, "Using the Canadian Censuses of 1852 and 1881 for Automatic Data Linkage: A Case Study of Intergenerational Social Mobility," in *Population*

increasingly brings computing science innovations together with traditional historical and literary sources.[50] This kind of record linking will make Canada the first country in the world with representative linked samples of comprehensive size at every ten-year interval between 1852 and 1921. To suggest only a few of the many possibilities for research opened up by these data, we will be able to follow tens of thousands of infants enumerated in 1852 until they reach advanced ages in 1921; analyze how their lifetime trajectories differed by social class, ethnicity, and location; compare their experience with that of their fathers; and trace the influence of parental circumstances on children and grandchildren. There are so many other opportunities, some of which we cannot yet even imagine!

TCP/Lpc data also overcomes the limitation of samples that are too small for analysis and statistical inference about important ethnic, historical, and geographic diversities. We will be able to study those individuals residing in particular regions or community types (coastal fishing towns, villages on the northern forest frontier, and so on) as well as people at various life course stages (for example, losing a spouse, leaving the parental home, marrying outside of one's ethnicity or faith). We will be able to undertake research on specific First Nations, people who migrated from particular locations and other groups that too often are overlooked. Moreover, TCP/Lpc will enable regionally differentiated comparisons of Irish Catholics versus Irish Protestants, francophones versus anglophones, First Nations who encountered Europeans in the seventeenth century versus those who did not meet whites until after 1900, to name a few. These and other research initiatives will explore the spatial and social diversities that underpin Canada's broad transformations from 1852 to 1921 with unprecedented precision.

CONCLUSION

The ongoing innovation of new digital resources, and the digitization of already available sources, presents enormous opportunities for the libraries, archives, and other public institutions that support Canadian history. With the opportunities come challenges including, as always, cost. The creation of databases is costly, as is the preparation of appropriate documentation and the institutions needed to preserve and disseminate data. Perhaps the greatest challenge is the effective preservation and ongoing dissemination of the data. Canadian institutional commitments in this area are surprisingly feeble. No dedicated budget is available to researchers or to professional associations for this purpose. Many individuals are independently digitizing source materials but without the resources to do it properly, without the coordination that would prevent, or at least minimize, a duplication of efforts, and without the means to guarantee

Reconstruction, ed. Bloothoft et al., *Population Reconstruction*, 243–61; Hélène Vézina et al., "The Linkage of Microcensus Data and Vital Records: An Assessment of Results on Quebec Historical Population Data (1852–1911)," *Historical Methods* 51 (2018): 1–16.

50 Chad Gaffield, "Words, Words, Words: How the Digital Humanities Are Integrating Diverse Research Fields to Study People," *Annual Review of Statistics and Its Application* 5 (2018): 119–39.

adequate preservation and continued dissemination.[51] Nor is there consensus among historians and fellow travellers in the social sciences about the design of metadata and other cross-disciplinary standardizations that would facilitate a wider use of resources created by a single individual for a particular research project.

An effective strategy for the long-term accumulation and use of historical databases is waiting on infrastructure and resources that are not yet in sight.[52] We have yet to see any coherent plan or the government support needed to sustain such infrastructure. Digitization of the historical censuses has received more support than most other sources. And, yet, the census is just one element, albeit an important contributor to the twenty-first-century "dilemma of plenitude."[53] New sources inevitably challenge the tools and skills of historians. Traditional research strategies and tool sets are increasingly inadequate to make sense of the burgeoning billions of pages of qualitative as well as numeric data now accessible in digital form. Other disciplines already have at least some of the methodological expertise. To a considerable extent, we can anticipate how geographers, sociologists, economists, and data scientists will use the new data. But what about Canadian historians? Will they acquire the skills needed to make use of big historical data?[54] Or will they leave the field to colleagues in other disciplines?

We believe that a modest degree of quantitative sophistication is within the reach of most historians. Some minimal training in quantitative methods, currently absent in most undergraduate and even graduate programs, would equip the next generation of historians to realize the research potential of new historical resources. This is not to reduce history to social science with old data. Historians will continue to ask their own questions and develop their own understandings using a broad diversity of sources. Increasingly, however, this

51 Ian Milligan makes a similar point about digital photographs. See Ian Milligan, "We Are All Digital Now: Digital Photography and the Reshaping of Historical Practice," in this issue.

52 There continues to be discussion of the challenge through the Portage network of the Canadian Association of Research Libraries (Portage Network, https://portagenetwork.ca) and the research funding councils (see Tri-Agency Statement of Principles on Digital Data Management, http://www.science.gc.ca/eic/site/063.nsf/eng/h_83F7624E.html?OpenDocument).

53 The phrase "dilemma of plenitude" was first used in this context by Carolyn Strange, "Stories of Their Lives: The Historian and the Capital Case File," in *On the Case: Explorations in Social History*, ed. Franca Iacovetta and Wendy Mitchinson (Toronto: University of Toronto Press, 1998), 25–48.

54 John Lutz and Chad Gaffield argue that history departments need to modernize their teaching of disciplinary methods in order to prepare students to function effectively in the digital world. See John Lutz, "Riding the Horseless Carriage to the Computer Revolution: Teaching History in the Twenty-first Century," *Histoire sociale/Social History* 68 (2001): 427–35; Gaffield, "Clio and Computers." In a 2010 review of curriculum in Canadian history departments, we found only one quantitative methods course in fifty English Canadian universities and five courses in nine French Canadian universities. See Peter Baskerville and Kris Inwood, "Canada's Digital Heritage: The Nation's Historical Data in an International Context," *A Knowledge Synthesis Grant on the Digital Economy*, Final Report to Social Sciences and Humanities Research Council, 1 December 2010.

process will be assisted by some familiarity with methods already used in the social sciences, alongside other digital tools distinctive to the humanities. The most pressing challenge of big historical data, ironically, may be the methodological expertise of historians themselves.

ACKNOWLEDGEMENTS

We are grateful for useful comments from Chad Gaffield, Ian Milligan, and the editors and anonymous referees of this journal.

PETER BASKERVILLE is Research Chair in Western Canadian History at the University of Alberta. He has published numerous articles and books including *Unwilling Idlers: The Urban Unemployed and Their Families in Late Victorian Canada* (with Eric W. Sager, 1998), *Sites of Power: A Concise History of Ontario* (2005), *Household Counts: Canadian Households and Families in 1901* (edited with Eric W. Sager, University of Toronto Press, 2007), *A Silent Revolution?: Gender and Wealth in English Canada, 1860–1930* (2008), and *Lives in Transition: Longitudinal Research from Historical Sources* (edited with Kris Inwood, 2015).

KRIS INWOOD is a professor of economics and history at the University of Guelph. He has published in the economic and quantitative social history of Canada and the broader British world. Recent articles include "Atmospheric Pollution and Child Health in Late Nineteenth Century Britain," *Journal of Economic History* (with Roy Bailey and Timothy Hatton, 2018), "Interracial Marriages in Early Twentieth-century Cape Town," *History of the Family* (with Johan Fourie, 2019), and "Historical Variations in Incarceration Rates for Indigenous Women and Men in British Columbia," *Howard Journal of Criminal Justice* (forthcoming).

PETER BASKERVILLE est titulaire d'une chaire de recherche en histoire de l'Ouest canadien à l'Université d'Alberta. Il a publié de nombreux articles et ouvrages, dont *Unwilling Idlers: The Urban Unemployed and Their Families in Late Victorian Canada* (avec Eric W. Sager, 1998), *Sites of Power: A Concise History of Ontario* (2005), *Household Counts: Canadian Households and Families in 1901* (codirigé avec Eric W. Sager, 2007), *A Silent Revolution? Gender and Wealth in English Canada, 1860–1930* (2008), et *Lives in Transition: Longitudinal Research from Historical Sources* (codirigé avec Kris Inwood, 2015).

KRIS INWOOD est professeur d'économique et d'histoire à l'Université de Guelph. Ses publications portent sur les aspects économiques et quantitatifs de l'histoire sociale du Canada et du monde britannique en général. Parmi ses articles récents figurent « Atmospheric Pollution and Child Health in Late Nineteenth Century Britain », *Journal of Economic History* (avec Roy Bailey et Timothy Hatton, 2018), « Interracial Marriages in Early Twentieth century Cape Town », *History of the Family* (avec Johan Fourie, 2019), et « Historical Variations in Incarceration Rates for Indigenous Women and Men in British Columbia », *Howard Journal of Criminal Justice* (à paraître).