The Promise and Problems of Quantitative Evidence in Canadian History

KRIS INWOOD*

CANADIAN HISTORIANS have acquired a taste for numbers. An examination of nine Canadian history journals indicates that use of quantitative information has increased considerably during the last two decades. A large proportion of scholarly publications now have some significant quantitative component even though in many cases much of the argumentation and evidence is qualitative.

One implication of the change is that historians are routinely collecting and storing in spreadsheets and data base programs information that may be useful to other researchers. The change in historical practice creates a number of challenges for Canadian archives, universities, journals, granting agencies, and historians themselves. The underlying problem is the lack of consensus about which data can or ought to be made available in a semi-public way, and how this should be done. We lack institutional arrangements and, more fundamentally, a cultural commitment to encourage the sharing of data. A need exists for concerted action on the part of academic institutions to ensure the preservation and diffusion of data made machine-readable by historians.

I begin with a simple content analysis of recent publications in Canadian history in order to demonstrate the relatively wide use now being made of quantitative information. Identification of articles with a significant quantitative component is complicated insofar as some papers rest almost entirely on quantitative sources even though few, if any, numbers are visible. Underlying this ambiguity lies a fundamental problem of distinguishing the implicit from explicit use of numbers.[1] Another pitfall is created by those

* Kris Inwood is an associate professor in the Department of Economics at the University of Guelph. The author is grateful for discussions with Gordon Darroch, Elizabeth Ewan, José Igartua, Doug McCalla, Jim Miller, Richard Reid, Jamie Snell, and participants at the 1993 meetings of the Ontario Archivists Association and the Canadian Historical Association. A research award from the Social Science and Humanities Research Council assisted in the preparation of this paper.

1 Peter Laslett, *Family Life and Illicit Love in Earlier Generations: Essays in Historical Sociology* (Cambridge: Cambridge University Press, 1977), p. 7.

Table 1 Total Number of Canadian History Articles in Selected Journals and Proportion of Articles with Quantitative Content

	1969 –1972	1973 –1976	1977 –1980	1981 –1984	1985 –1988	1989 –1992	All Articles 1969–1992
CHR	0.18	0.21	0.21	0.23	0.19	0.33	279
HP/JCHA*	0.16	0.28	0.21	0.36	0.54	0.56	207
RHAF	0.28	0.38	0.32	0.37	0.52	0.51	372
OH	0.18	0.08	0.10	0.29	0.19	0.22	340
ACAD	0.10	0.17	0.33	0.32	0.50	0.54	201
BCS**			0.34	0.35	0.47	0.50	160
HS/SH*	0.61	0.59	0.53	0.56	0.73	0.74	237
LBT		0.17	0.21	0.43	0.47	0.53	166
CPRH			0.67	0.71	0.65	0.64	78
Overall average	0.24	0.26	0.30	0.38	0.46	0.48	
Total articles	264	287	361	389	380	359	2,040

Abbreviations:
CHR *Canadian Historical Review*
HP/JCHA *Historical Papers* (Journal of the CHA)
RHAF *Revue de l'histoire de l'Amérique française*
OH *Ontario History*
ACAD *Acadiensis*
BCS *B.C. Studies*
HS/SH *Histoire sociale/Social History*
LBT *Labour/le Travail(leur)*
Definitions and criteria are described in the text. The quantitative proportion is the number of articles judged to have quantitative content divided by the total number of articles.
* Only the Canadian articles are examined.
** Only the historical articles are examined.

papers, thankfully limited in number, that contain graphs and tables with little connection to the central point of the article.

A relatively simple approach to content analysis, which does not mask the crudeness of the exercise, is appropriate. I classify articles according to the apparent significance of numbers, if any, in the text, tables, graphs, or maps. The presence of a single table or a few text references to prices or population does not suffice to indicate quantitative content. Rather, the papers identified below are those employing data in a visible and significant manner.[2] Undoubtedly I miss some quantitative-based discussion that does not report the data in tables and graphs. For these and other reasons, the

2 The concept of papers having a quantitative component is broader than that employed in previous surveys; see Peter George and Ernest Oksanen, "Recent Developments in the Quantification of Canadian Economic History", *Histoire sociale/Social History*, no. 4 (November 1969), pp. 76–98; and José Igartua, ed., "Historical Databases: The Canadian Experience", a round table discussion, *Histoire sociale/Social History*, vol. XXI, no. 42 (November 1988), pp. 283–318.

exercise provides no more than an approximate indication of the extent to which Canadian historians now use quantitative evidence.

The nine journals examined here have been central to the writing of Canadian history during the last 25 years. They are the *Canadian Historical Review*, *Journal of the Canadian Historical Association* (formerly *Historical Papers*), *Acadiensis*, *Revue de l'histoire de l'Amérique française*, *Ontario History*, *B.C. Studies*, and three thematic journals, *Histoire sociale/Social History*, *Labour/le Travail*, and *Canadian Papers in Rural History*. I consider all papers, presidential addresses, and research notes on Canadian history topics. The restriction of coverage to papers that are clearly historical in nature results in the exclusion of a number of papers in *B.C. Studies*, just as the Canadian focus excludes many papers in *Histoire sociale/Social History* and *Journal of the Canadian Historical Association*. Short as well as long pieces are included, although editors' notes, bibliographies, review articles, and comments, critiques, and replies are excluded (unless the reply presents information clearly absent from the original article).

I summarize the result of this simple content analysis in Table 1. Nearly one-half of all Canadian history articles now have quantitative content. This figure represents a significant increase from an average level of one-quarter circa 1970. The quantitative proportion today undoubtedly would be even higher if we were to include Canadian papers appearing in international journals and the journals of allied disciplines.[3] Note also that I have not considered monographs or papers in published collections.

The individual journals differ in various ways. Thematic journals, for example, tend to be more quantitative than the generalist ones. Change has come more slowly to the *Canadian Historical Review* than to the *Journal of the Canadian Historical Association*. One of the regional journals, *Ontario History*, consistently has been the least quantitative journal of those examined here. At the other extreme, *Histoire sociale/Social History* began as a heavily quantitative journal and has become more so as social historians examine groups and classes that are voiceless in more traditional sources.

It would be foolish to suggest that any one pattern is better or more appropriate than others. Quantitative evidence is more helpful for some topics and methods than others for a variety of reasons, including the idiosyncrasies of source survival. Moreover, the strength of Canadian history journal publishing lies precisely in its diversity; there is good reason to welcome the distinctive identities established by individual journals. This survey of journal content shows, however, that Canadian historians now use quantitative evidence much more than they did 25 years ago. Such evidence is used even in traditionally non-quantitative fields such as religious his-

3 I am thinking of the *American Historical Review*, *Journal of Family History*, *Agricultural History*, *Journal of Economic History*, *The Canadian Journal of Economics*, *Canadian Review of Sociology and Anthropology*, *Canadian Geographer*, and so on.

tory[4] and political history[5]. The use of data is no longer confined to a small number of 'number-crunching' historians. The widening use of quantitative information brings with it considerable potential as well as challenges with respect to data accessibility and sharing.

One advantage of quantitative information is the ability to store and analyze it in a standardized, machine-readable form, which is relatively easy to share among various users. Consequently we have entered a research environment in which information gathered in the course of one project may be useful to others. The most dramatic examples of this possibility are research projects in other countries that have succeeded in attracting the necessary financial support in part because one promised product of the research agenda was to generate data sets useful to others. The result has been research publications that would not otherwise have been possible both from the initial project and from subsequent researchers.

Canadian history has seen few projects that promise to make data available for the use of others. The recent release of 1891 and 1901 census manuscripts, for example, has attracted little interest in the generation of a user sample to serve the long-term needs of the scholarly community as well as the immediate research needs of a few individuals. Instead, we have several small projects, valuable in themselves, but conducted with distinctive data entry principles that may or may not permit consolidation into a carefully structured sample such as that now available for the 1901 United States census.

It is understandable that Canadian scholars eschew the co-ordination and management difficulties that inevitably accompany a project large enough to produce data sets as a byproduct of the research endeavour. It is perhaps more surprising to find so little evidence of data sharing of any sort in Canada. My survey of journal content encountered few articles employing data collected by someone else, in spite of the significant increase in the use of quantitative materials. A remarkably wide range of sources is now being transferred into machine-readable form but very little is ever used by anyone except the original researcher, although exceptions arise from the important research networks in Quebec.[6]

4 For examples, see Jean Roy, "Le clergé nicolétain, 1885–1904 : aspects sociographiques", *Revue de l'histoire de l'Amérique française* (décembre 1981), pp. 383–395; and Christine Hudon, "Carrières et vie matérielle du clergé du Richelieu-Yamaska (1790–1840)", *Revue de l'histoire de l'Amérique française* (printemps 1992), pp. 573–594.

5 For examples, see D. A. Muise, "Parties and Constituencies: Federal Elections in Nova Scotia 1867–1896", Canadian Historical Association, *Historical Papers* (1971), pp. 183–202; and Gail Campbell, " 'Smashers' and 'Rummies': Voters and the Rise of Parties in Charlotte County, New Brunswick, 1846–1857", Canadian Historical Association, *Historical Papers* (1986), pp. 86–116.

6 I am thinking of the Programme de recherche en démographie historique (PRDH) at the Université de Montréal and the Centre interuniversitaire de recherches sur les populations (SOREP). Another initiative that seems likely to find wide use is the Ontario Genealogical Society's machine-readable

Of course, much of the quantitative material identified in the survey of journal content does not lend itself to sharing. In many cases the original source would have to be re-examined at length by other researchers, even if the spreadsheet or data base file were available to them. In other cases, the transcription process is very complex and embodies judgements so complicated that the kind of documentation needed by others would be exceedingly time-consuming to produce. Other data sets are so particular that the likelihood of interest by others is too small to justify the cost of archival preservation.

It is possible to imagine numerous reasons why one researcher might use quantitative materials and leave no legacy of machine-readable data sets useable by others. Precisely for that reason, however, historians must begin to develop principles to distinguish data sets that deserve the full archival and preservation approach. It may be difficult to secure consensus around an elaborate code of conduct, but agreement should be realizable on the general principle that the treatment of data sets is a matter of professional ethics and on certain summary guidelines.

One indication that we do not yet have an academic culture that puts a positive value on the sharing of data is our apparent willingness to tolerate researchers who adopt a narrowly proprietorial attitude towards "their" data. A number of situations have existed during the past five years in which Canadian professors have declined to allow other academics access to their research data, whether to retain exclusive use, to avoid criticism of published articles, or to raise revenue through the sale of data. Each is clearly undesirable. Extreme examples, however, divert attention from the lost opportunities of many small projects, which typically make no arrangements for public access to data sets prepared and used in the course of research. Very few researchers deposit their data sets in a public institution or even let it be known that the data are available for others to use. Not surprisingly, there has been some duplication of effort at public expense as a result of historians independently entering the same data. No doubt there are many examples in which public accessibility is not an issue, as is suggested above. Nevertheless, the only effective way to determine whether a data set may be useful to another is to bring it into the public domain in a way that is visible and truly accessible.

Full accessibility is impossible unless minimal and known standards are employed in the preparation and documentation of data. This is more time-consuming and contentious than might appear at first glance. We have at least one example of a research team having fallen apart over the issue of

---

version of the 1871 census schedule 1, supervised by Bruce Elliott in consultation with staff at the National Archives of Canada. The creation of this data set coincided with the preparation of Professor Elliott's *Irish Migrants in the Canadas: A New Approach* (Kingston and Montreal: McGill-Queen's University Press, 1988).

data quality, resulting in the circulation of two distinctly different versions of the same manuscript data. The lesson of the story is that we must strengthen the incentives for researchers to take care in the design of their samples, the structuring of records, and the preparation of documentation. The appropriate place to begin teaching these skills is coursework in graduate history programs. A survey of this aspect of graduate education, possibly conducted by the Canadian Historical Association, would provide useful information on this point.

The considerable time needed to correct, document, and organize a data set in a way that others will find useful carries another implication. It is essential that university appointments, tenure and promotion committees, and adjudication committees of the Social Sciences and Humanities Research Council (SSHRC) recognize the time and expense incurred in the proper preparation of data sets. Recognition of and respect for appropriate standards of data preparation also imply that university and SSHRC committees acquire the expertise needed to assess quality in this aspect of research activity.

A profitable comparison might be drawn with the literary scholar who prepares a critical edition of a text or document. The latter is easily recognized as a valuable scholarly contribution, which is amenable to evaluation using agreed-upon methods of peer review. In principle, there is little reason to treat a large data set in any other manner. Journals, for example, might begin to review data sets in a manner comparable to the appraisal of critical editions, document collections, and other complex research tools.

A second journal-related issue is the development of standards for citation of data bases used in scholarly publications. The Canadian Historical Association's Committee on Computing and History at its 1993 meeting considered data base citation guidelines which José Igartua has proposed to *Revue de l'histoire de l'Amérique française* and *Histoire sociale*. It is likely that these and other journals will soon adopt these or similar guidelines that indicate clearly the original archival source as well as the source of the machine-readable version.

A thornier issue is the establishment of procedures regarding data used in published articles. The journals in most social science disciplines have begun to insist that contributors send their data along with the submission or agree to make it available to anyone who might wish to contest the results of a published article. The readers of historical journals arguably should have some opportunity to persuade themselves that published reports of difficult-to-access data are in fact correct. Journal procedures along these lines would help to dispel the occasional scepticism with which quantitative research is received.

A need for policy initiatives by granting agencies (most obviously SSHRC) and universities becomes apparent if we remember that academic careers advance on the basis of research publications rather than the responsible preparation of data sets. Even if an individual historian is inclined to be responsible with respect to data, there is absolutely no incentive to divert time and money from a publication project in order to document and make available data to other scholars who are, in a sense, competitors. It is essential that our research institutions strengthen the incentives facing the individual scholar. The difficult challenge, however, is to do so in a way that supports rather than undermines the individual researchers who, after all, are responsible for the generation of machine-readable data sets. The most obvious strategy is to provide appropriate funds for data base development, rewarding the researcher with suitable recognition and ensuring some reasonable but clearly specified period of exclusive use.

Some historians have not yet brought their data into the public domain because it has not been obvious how to do so. Fortunately, we are beginning to see the emergence of a new institution, the data archive, which will ensure the survival of data for the use of researchers other than the creator of the data set. The purpose of a data archive is to store, maintain, document, and make available electronic data sets.[7] Just as libraries have developed data library divisions, archives are beginning to establish data archives. As yet, however, few Canadian archives even within universities have made the necessary institutional adjustment, in part because start-up funds are scarce in an era of limited public sector budgets. Historians can only hope that reductions in the cost of computer hardware and software will permit initiatives in the near future to establish one central data repository or a network of individual repositories with the momentum to become a major presence in our national research infrastructure.

One small but important step would be to incorporate into the users' agreement at every public archive a clear commitment to document and deposit any machine-readable files created from archival collections. This step is needed because Canadian law permits a declaration of copyright simply by transferring information into a machine-readable format. The result is to remove such material from the purview of archives and, effectively, from the public domain. It is astonishing, for example, that machine-readable versions of historical census data are not in the public domain even though public funding made possible collection of the original data, their preservation until the present day, and their transfer into a machine-readable format. It may be possible to amend this law for the special purpose of protecting our national heritage, but until that happens archives can try to limit the damage through an appropriately worded user agreement.[8]

7 See Gordon Darroch and Sue Gavrel, "Preserving Historical Databases and Facing Technical Change: Common Issues for Social Historians and Archivists", *Archivaria*, 34 (Summer 1992), pp. 288–297; and Kris Inwood and Richard Reid, "The Challenge to Archival Practice of Quantification in Canadian History", *Archivaria*, forthcoming.

8 Strictly speaking, the purpose of a legislative amendment might be to clarify rather than change the law, since there is considerable ambiguity at the present time. Difficulties arise with researchers who exploit the ambiguity by claiming copyright for data sets and construing its implications very

Archives, the universities, journals, and granting agencies each can adopt policies reinforcing our collective commitment to the accumulation of a shared pool of machine-readable information and hence to an efficient use of scarce research funding. Of course, effective policy in this area must be guided by a recognition that historical data differ from those collected by natural and most social scientists. Not all historical data fall under the broad rubric of public accessibility. Historical data typically are more ambiguous and difficult to interpret because they are not collected under controlled or carefully selected conditions. For these and other reasons historians must adapt rather than borrow procedures and practices used in other disciplines.[9] We may hope that the various institutions will facilitate this process with policies to encourage the creation of useful new data bases and their diffusion within the research community.

---

broadly. The difficulty of interpreting copyright law as it applies to data bases is mentioned by Rhonda Birenbaum, "Copying Right", *University Affairs* (August–September 1993), pp. 6–8.

9 An explicit reference to historical data in Appendices D or E of the *SSHRC Grants: Guide to Applicants* would help to define the ways in which historical data are distinctive and yet subject to the same general concern for accessibility as applies to the social sciences.