

The Challenge to Archival Practice of Quantification in Canadian History

by KRIS INWOOD and RICHARD REID*

The use of quantitative evidence in Canadian history has increased dramatically during the last twenty-five years. While archival repositories hold much of the original (non-electronic) data, many historians have found it useful to transfer small or even large portions of various archival collections into machine-readable format. This movement of information into the electronic arena threatens archivists with the loss of their traditional role as custodians of the basic evidence on which history is written. Historians are likely to support archivists in their response to this challenge, particularly if the response involves some clarification of responsibility for machine-readable datasets. Archival institutions are in a position to help maintain and ensure access to electronic data originating in archival sources.

The increase in quantitative content in Canadian historical writing reflects the gradual fading of a long-familiar division between quantitative and non-quantitative methodologies. A common image of the quantitative historian is of someone trained in social science rather than history, or of someone whose career has been dedicated to exploitation of quantitative sources. In fact, few authors rely exclusively on quantitative sources. Today it is common for an individual historian to consult a wide range of both quantitative and qualitative sources. These authors would not regard themselves as quantitative historians, even though they use quantitative data to set the context for their work or to illuminate one aspect of a larger topic. While many papers have some significant quantitative component, much of their argumentation and evidence is qualitative. Such articles appear in traditionally non-quantitative fields such as religious history¹ or political history,² as well as in fields that are heavily quantitative in nature, such as demographic or economic history. Much of the use of quantitative materials reflects the efforts of social historians to give voice to people who leave no other records.

In this paper we monitor the use of quantitative sources by historians in order that we may illustrate more clearly the nature of the challenge facing archivists today. The breakdown of old methodological paradigms has complicated our task. We have made no attempt to isolate a body of literature that constitutes "quantitative history." Rather, it seemed more useful to identify articles that had a significant quantitative component. This approach is broader than that employed in previous surveys, but it is more in keeping with the growing tendency to incorporate qualitative and quantitative sources simultaneously.³ Our definition of "article" encompassed all papers, presidential addresses, and research notes; we restricted our attention to papers on Canadian topics. Length was not a criterion—short as well as long pieces were included. We excluded editors' notes, bibliographies, review articles, comments, critiques, and replies (unless the latter presented information clearly absent from the original article).

The basic criterion used to classify the content of articles was the presence of numbers, either in the text of the article or presented in tabular form. Of course, application of the criterion required some judgement. A paper with one small table or a table in the notes does not qualify as quantitative unless the information is central to the argument of the paper. Similarly, a few text references to prices or population do not suffice to indicate quantitative content, although an extended quantitative description would do so. Undoubtedly we missed some text discussion about or based on quantitative evidence because it was less obvious at a glance than tabular presentation: the qualitative presentation of quantitative analysis has become common.⁴ Our procedure is only one way to classify the quantitative aspects of journal content; it does have the virtue, however, of indicating in an approximate way the extent to which Canadian historians use quantitative evidence.

We examined five of the journals that have been central to the writing of Canadian history during the last twenty-five years. The *Canadian Historical Review (CHR)* is published by the University of Toronto Press, *Histoire sociale/Social History (HS)* is issued by the University of Ottawa, and the Canadian Historical Association publishes the *Journal of the Canadian Historical Association/Revue de la Société historique du Canada (JCHA/RSHA - formerly Historical Papers)*.⁵ *Acadiensis (AC)*, issued by the University of New Brunswick, and *Revue d'histoire de l'amérique française (RHAF)*, published by L'Institut d'histoire de l'amérique française, were also selected as representative regional journals.⁶ A listing of the articles judged to have quantitative content is available from the authors on request.

A summary description of the content classification in **Table 1** indicates that the proportion of articles with quantitative content has roughly doubled since the late 1960s.⁷ The journals differ enormously, of course. During the recent past, quantitative evidence appears in about one-third of the articles in *CHR*, one-half of the articles in *AC*, *JCHA*, and *RHAF*, and three-quarters of the articles in *HS*. The journals also differ in the path by which they have arrived at their current practice.

Articles in the *CHR*, *JCHA*, and *AC* made little use of quantitative sources at the beginning of our period. However, since 1970 each journal has changed—albeit in different ways. Until recently, the *CHR* maintained a relatively low level of quantitative content, and it remains the least quantitative journal—although it has been changing rapidly in recent years. *AC* increased its quantitative share during the mid-1970s and again during the late 1980s; one-half of its articles now have quantitative content. The *JCHA* experienced a particularly dramatic increase in quantitative content during the 1980s. This journal provides the most effective indicator of practice among young scholars, since the articles are selected from papers presented at the association's annual meeting. The quantitative share of *JCHA* articles has not yet overtaken that of *HS*, which throughout the period has been the most quantitative Canadian journal. The heavily quantitative nature of *HS* no doubt reflects the attempt by social historians to examine groups and classes who are voiceless in more traditional historical sources. *RHAF* is noteworthy in that its experience closely matches the collective average of the five journals in level of quantification and change over time.

Table 1

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

		AC	RHAF	CHR	HS	JCHA
Total Number of Articles Published		201	372	279	237	207
Quantitative Proportion	1969-72	.10	.28	.18	.61	.16
	1973-76	.17	.38	.21	.59	.28
	1977-80	.33	.32	.21	.53	.21
	1981-84	.32	.37	.23	.56	.36
	1985-88	.50	.52	.19	.73	.54
	1989-92	.54	.51	.33	.74	.56

Abbreviations:

AC *Acadiensis*⁸RHAF *Revue d'histoire de l'amerique francaise*CHR *Canadian Historical Review*HS *Histoire sociale/Social History*JCHA *Journal of the Canadian Historical Association/Revue de la Société historique du Canada (formally Historical Papers)*

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.

The individual identities of the different journals make it difficult to speak about Canadian historical writing as a whole. Nevertheless, three general observations are possible:

- (1) The proportion of historical writing using quantitative evidence has roughly doubled during the past twenty-five years; much of the change occurred during the 1980s.⁹
- (2) The period of rapid change may already have ended; extrapolation from the recent past gives no reason to expect future increases in quantification (with the possible exception of the *CHR*).¹⁰ On the other hand, there is no reason to expect a decrease, given the trend of *JCHA* content, which is heavily influenced by the activity of younger scholars.
- (3) Overall, about one-half of Canadian historical writing now relies upon some degree of quantification. The proportion undoubtedly would be higher if we were to include Canadian articles appearing in international journals and the journals of allied disciplines.¹¹

It would be incorrect, of course, to equate quantitative evidence with electronic or machine-readable data, since it is now common to store texts, image, and qualitative material of all sorts in machine-readable form.¹² Moreover, paper transcription and hand calculation continue to provide an effective basis for the analysis and presentation of small amounts of data. The cost advantage of machine-readable technology lies largely with large datasets and/or complicated calculations.¹³ Nevertheless, a good deal of quantitative information is stored and analyzed electronically. It is perhaps not surprising, therefore, that quantification has led to the widespread conversion of data in archival collections into machine-readable form. Some researchers assert copyright over the electronic version of information simply by virtue of their having converted into machine-readable format. In such cases the movement of data into the electronic

arena removes it from the purview of archives and, effectively, from the public domain. Some of these collections, census manuscripts being one example, comprise a significant part of our national heritage.¹⁴

The increasing importance of electronic data, whether it is new information existing only in electronic form or machine-readable datasets transcribed from historical documents, directs attention to the concept of a data archive—the purpose of which is to store, maintain, document, and make available to users electronic datasets.¹⁵ Just as libraries have developed data library divisions, archives now must begin to think about taking responsibility for data archives.

The emergence of data archives in Canada has been hindered by the financial barrier imposed by start-up and maintenance costs. In fact, late entry confers some advantages. Today, hardware is much less expensive, software is easier to use, and both are more powerful, than was the case five years ago. More sophisticated technology such as long-distance networking is now available, and already has begun to serve archivists well. The pace of change in information technology and historical practice dictates that archival institutions should begin to plan their presence in the field of electronic information if they wish to preserve their role as specialists in the storage and care of unpublished information.

Historians have a particular reason to support such an initiative because, in its absence, they have become the custodians of archival datasets. Numerous researchers have transcribed into machine-readable form information found in archival collections. In most cases, a government department or other *public* agency was responsible for collection of the original data, and *public* archives have maintained the information until the present day. The individual researcher typically receives a university salary or research-time stipend during the transcription project. *Public* funds typically cover most auxiliary costs, including that of research assistants, who often do most of the work. Given the strong element of public funding, it would seem self-evident that these datasets should be considered to be in the public domain, and that the researcher should be obliged to undertake transcription in a way that observes appropriate standards of data integrity and makes them available for others to use at little or no cost.

Unfortunately, it almost never happens that way.¹⁶ Many researchers do not observe minimal standards of data integrity, and/or fail to document the basic characteristics of the dataset in a way that would facilitate its use by others. Even when an effort is made, it is often inadequate. There is already one example of a research team having fallen apart over the issue of data quality—resulting in the circulation of two distinctly different versions of the same manuscript data. Very few researchers deposit their datasets in public institutions or even let it be known that the data are available for others to use. Not surprisingly, there is now considerable duplication as a result of historians independently entering the same data.

The challenge to archival practice is rather more serious than might be imagined. We must not think only about the handful of Canadian historians who use big datasets and little else in the way of sources. These authors are few in number; they tend to rely heavily on a few, relatively important sources. Equally significant, however, has been the largely unnoticed increase in the use of quantitative materials by traditional historians. These scholars use a remarkably wide range of sources, some of which are transferred into machine-readable form.

The underlying problem is that historians are not well suited to the task that they have inherited. No doctoral programme in history offers instruction in the preservation of machine-readable data.¹⁷ More fundamentally, the craft of history is a solitary one; most good historians are individualistic and even entrepreneurial in nature. Moreover, academic careers advance on the basis of research publications rather than the responsible preparation of datasets. 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 seems clear that historians are not the appropriate group to acquire responsibility for machine-readable datasets. Some of the same objections apply to historians acting collectively through their professional association. A role for archival institutions, on the other hand, is implied by their traditional responsibility to maintain the basic sources with which history is written. It seems a natural extension of this role to include machine-readable datasets, which have their origin in archival collections.

One other institution is relevant to this discussion. Many university campuses already have a machine-readable data library unit dedicated to the maintenance of datasets—most of which have been published by national governments.¹⁸ In some cases, these units are formally part of the university library system; others are located within departments, faculties, and other academic entities. Their existence offers important opportunities for collaboration among archivists, librarians, and academic staff.

An even broader institutional collaboration is possible with recent improvements in data transfer and storage capacity—which now make it feasible to store in one location all machine-readable datasets of possible interest to historians and other users of Canadian archives. There is not yet a machine-readable “public use sample” of the 1901 Canadian manuscript census, for example; a master copy in one location might be accessed within minutes or even seconds by users throughout Canada. There would be no need for the Public Archives of Nova Scotia to maintain a copy of the Maritime observations, because Internet would allow the relevant data to be accessed and transferred quickly, as they are needed, to Halifax.

Ease of transfer implies that a single archive of machine-readable data would have the capacity to serve all Canadian archival repositories. Presumably such a facility would function along the lines of the Inter-University Consortium for Political and Social Research (ICPSR) at Ann Arbor, or the University of Essex/ESRC Data Archive, which has an active History Data Unit.¹⁹ The advance of network technology also makes it possible that a future Canadian archive of machine-readable data would be a network, rather than a single large computer maintained by a particular institution. Each of the nodes or archival institutions participating in the network would absorb costs in the form of hardware, software, staff time to develop common standards and procedures, and, most of all, staff time for training. One advantage of the network concept, however, is that these costs would be lower than would be the case if each institution entered the electronic arena independently.

Another issue which may attract discussion is the *Copyright Act*, which permits the machine-readable version of historical census materials to be removed from the public domain. The Association of Canadian Archivists, perhaps in conjunction with the Canadian Historical Association, might want to take a position on this issue.

Finally, a small but easy step would be to incorporate into the agreement signed by users at every public archive a clear commitment concerning the eventual status and disposition of machine-readable files created from archival collections.

Any or even all of these steps, by themselves, will not resolve the various issues created by the largely unregulated transfer of archival materials to electronic media. The Canadian Historical Association, the Social Sciences and Humanities Research Council of Canada, and Canadian universities are also in a position to undertake useful measures. Regardless of action taken by others, however, archivists—through their professional associations and institutional employers—will be central participants in the collective response to a process which has its roots in the use of quantitative sources by Canadian historians.

Notes

- * An earlier version of this paper was presented to the 1993 meeting of the Archives Association of Ontario. Ideas that came forward during the discussion in Guelph led to the writing of this paper. Helpful comments from Jamie Snell are gratefully acknowledged. A research award from the Social Science and Humanities Research Council of Canada assisted in the preparation of this paper.
- 1 For examples see Jean Roy, "Le clergé nicolétain, 1885-1904: aspects sociographiques," *RHAF* (décembre 1981), pp. 383-95, and Christine Hudon, "Carrières et vie matérielle du clergé du Richelieu-Yamaska (1790-1840)," *RHAF* (Printemps 1992), pp. 573-94.
 - 2 For examples see D.A. Muise, "Parties and Constituencies: Federal Elections in Nova Scotia 1867-1896," *JCHA* (1971), pp. 183-202, and Gail Campbell, "Smashers and Rummies: Voters and the Rise of Parties in Charlotte County, New Brunswick, 1846-1857," *JCHA* (1986), pp. 86-116.
 - 3 Peter George and Ernest Oksanen, "Recent Developments in the Quantification of Canadian Economic History," *SHHS* (November 1969), pp. 76-98, and José Igartua, ed., "Historical Databases: The Canadian Experience," *SHHS* (November 1988), pp. 283-318.
 - 4 On the other hand, some articles may have been included on the basis of tabular data that was peripheral to the principal thrust of the argument.
 - 5 We exclude *Labour/Le travail*, *Canadian Papers in Rural History*, and *Canadian Papers in Business History* because they appeared for the first time during the period under study.
 - 6 *Acadiensis* and *Revue d'histoire de l'Amérique française* have been particularly successful at attracting an extra-regional audience. We recognize as well the importance of other journals such as *BC Studies*, *Alberta History*, *Saskatchewan History*, *Manitoba History*, *Prairie Forum*, *Nova Scotia Historical Review*, *La société historique acadienne*, etc.
 - 7 The initial date of 1969 reflects our organization of the data into four-year intervals and the emergence of *Histoire sociale* and *Acadiensis* during the late 1960s and early 1970s.
 - 8 *Acadiensis* did not begin publication until 1971.
 - 9 The comparison here is between the quantitative proportion in 1977-80 and that in 1989-92.
 - 10 The comparison is between the quantitative proportion in 1985-88 and that in 1989-92.
 - 11 We are thinking of international journals such as the *American Historical Review*, *Journal of Family History*, *Agricultural History*, *Journal of Economic History*, etc., and the journals of allied disciplines such as *The Canadian Journal of Economics*, *Canadian Review of Sociology and Anthropology*, *Canadian Geographer*, etc.
 - 12 Deian R. Hopkin, "Educational Technology and the Study of History: the British Experience," paper presented to the Annual Meeting of the Canadian Historical Association, Ottawa, June 1993. An interesting example in this genre is described by Thomas Martin, "The Electronic Databases on Classical Greece of the Perseus Project: Reactions to and from a Critical Mass of Data," paper presented to the Annual Meeting of the Canadian Historical Association, Ottawa, June 1993.
 - 13 It is worth remembering that machine-readable technology is now used to store and analyze text databases and other information traditionally seen as qualitative.
 - 14 Census materials were the subject of discussion at the Conference on the Use of Census Manuscript Information for Historical Research, University of Guelph, March 1993.
 - 15 Various issues surrounding machine-readable archives are discussed by David Bates, et al., "Promises, Promises: The Problems and Prospects of Machine-Readable Archives," (York Institute for Social Research, 1986), and Gordon Darroch and Sue Gavrel, "Preserving Historical Databases and Facing Technical Change: Common Issues for Social Historians and Archivists," *Archivaria* 34 (Summer 1992), pp. 288-97.
 - 16 One of the few projects that does fit the desirable pattern is the Ontario Genealogical Society's machine-readable version of the 1871 census schedule 1. This project was supervised by Bruce Elliott of Carleton University, in consultation with staff at the National Archives of Canada.
 - 17 Indeed, few Canadian history departments offer training in any aspect of quantitative history.
 - 18 There is, of course, uncertainty about the distinction between published and unpublished materials in electronic media. Consider, for example, a hypothetical household survey undertaken by Statistics Canada. Tabulations of the data are published in paper format and acquired by libraries. The original paper copies of the survey instruments remain in the hands of Statistics Canada, where they have an archival status (albeit an untouchable one because of confidentiality concerns). A machine-readable version of the raw data is identical to the archival material, except that individual household identification is removed. In that sense the machine-readable version is an archival resource; yet it is also available to other users, and hence may be construed as a publication.

- 19 Sheila Anderson, "The ESRC Data Archive as a Resource Centre for the Future," *History and Computing* 4 (1992), pp. 191-96. Comparable archives in other countries would include Zentral Archiv fur Empirische Sozialforschung (Germany), Norwegian Social Science Data Archive (Sweden), Belgian Archive for the Social Sciences (BASS), Archivio Datie Programmi per le Scienze Sociali (Italy), Banque de donnés social-politiques at the Institut d'études politiques (France), Steinmetz Arcives (The Netherlands), Danish Data Archives, WISDOM (Austria). The National Archives of the United States and Canada hold public records in the Center for Electronic Archives (US) and Government Archives Division (Canada).