

What Happened to Our Data? / Que sont nos données devenues?

Whose Data? Who Decides? An Editor's View

J.R. MILLER*

ONCE UPON A TIME there was a scholarly journal that unwittingly became embroiled in a dispute between two sets of researchers over data. A pair of relatively junior investigators submitted a comment upon a paper that had been prepared by a relatively senior duo of researchers, and which had been published in the journal. Complicating the case was the fact that the junior scholars' comment, which was critical of their seniors' published article, relied in part upon a set of data that had been collected by one of the original authors. When this author became aware of the comment, he objected to the use of what he regarded as "his" data by the other two. For their part, the would-be commentators believed that they had permission to use the data, which had been transmitted to one of them by the other senior author. However, the scholar under whose direction the disputed data set had been assembled said that he had not been consulted about access to "his" data set, had not given permission for access or use, and would not have agreed to providing access and permission to use the set if he had been asked. This imbroglio, which came to the attention of an editor after the proposed commentary had been submitted to be considered for publication, naturally caused no little problem.

One aspect of the dispute that ensued was the delay that it inflicted on the journal's adjudication process. It took almost seven months for the disagreement over use of the data set to be resolved; as a result, the adjudication process started about half a year later than it might have. A second problem that profoundly troubled the editors was the potential for finding themselves and the publisher of the journal the objects of litigation if they proceeded to consider the submission from the more junior scholars. Given the parlous financial condition in which most journals and their publishers find themselves in the 1990s, the editorial team was going to err, if it erred, on the side of caution. This "tilt" in the direction of legal prudence clearly gave an advantage to one side in the dispute between the

* J. R. Miller is co-editor of *The Canadian Historical Review*.

two teams of investigators. With cautious editors at the journal's helm, a delay in resolving the dispute over data simply meant that the work of the more junior team would be stalled.

The incident raised other issues, too, which are summarized in the title of these remarks. Whose data are they, anyway? In the event that there is a dispute over that question, who decides whose data they are? In the case to which I am referring, the member of the more senior team who had been responsible for the assembly of the disputed data set maintained that the data were "his". The junior investigator contended that, since the data had been collected with the assistance of a grant from a publicly supported funding agency, they ought to be available to all after a reasonable period in which the collectors would have an exclusive right to exploit them. Indeed, this disputant pointed out that the granting agency in question had a "policy" whereby such data were to be made available to other scholars after three years. This did not persuade the more senior researcher, who replied that he had contributed some of his own funds to support the data collection. These questions were not debated to a conclusion because the disputants eventually resolved their differences. However, the question "Whose data are they?" remains for the community of historical investigators in general.

Unfortunately, we in Canada do not have any agreed-upon referee to settle disputes over proprietorship of, access to, and right to use data. There are three potential arbiters, none of them logically more appealing than the other – and probably none of them willing. First is the community of historical researchers itself. It would be wonderful if something like the Canadian Historical Association or an appropriate subgroup could be empowered to deal with and settle such differences of opinion. However, the CHA has quite enough to do with what is already on its agenda. In any event, how would either the CHA or any other body of professionals enforce its rulings? What sanction could it apply? Could it bar recalcitrant data-hoarders from its annual meeting? Refuse to let them give papers? Prevent their attendance at the beer tent? Not likely.

A second, probably equally unlikely candidate for the role of reluctant Solomon is the granting agency. Granting agencies, however, have neither the personnel nor the "clout" to adjudicate disputes and enforce rulings. How often has the Social Sciences and Humanities Research Council of Canada been asked to enforce its "policy" that data collected with Council funds are to be made generally available after three years of exclusivity? How often has it responded to such requests? What have been the results? More to the point, is it reasonable to ask a public agency to take on a responsibility that belongs to the scholarly community itself?

Third, obviously, are the publishers as possible referees for disputes over data. By "publishers" I mean both journals and scholarly presses, though I recognize that sometimes the two are the same corporate entity. (*The Canadian Historical Review* is owned and published by the University of

Toronto Press, Inc.) In favour of having publishers charged with policing access to and use of data is the argument that they have leverage or “clout” with researchers. They can use access to their adjudication processes – and potentially to publication – to force those who assemble data to make them available to others for verification, for refutation, and for elaboration of the original results. Publishers, then, *can* carry out this task of deciding “Whose data are they?”

This brings us logically to the question of *how* publishers should do so. There are two options: an internal and an external method. The better is the latter, by which an agency such as the National Archives of Canada would serve as a repository for machine-readable data sets. It could establish or facilitate the establishment by others of uniform standards for data collecting and recording, it could guarantee safe storage, and it could disseminate the data to any legitimate investigator after an initial period of exclusive use by the collector. If such a system existed, then publishers of monographs and scholarly journals could require contributors who base their material upon quantitative data to deposit a copy of the data set with the authorized repository prior to submission of their material to the publisher. Until the publisher received proof of such deposit, it would not proceed with the adjudication of the material submitted. However, if this is the optimum arrangement, it is not one that is available to the research community in this country in 1993.

The unavailability of a central repository with uniform standards and procedures probably will force those publishers who choose to do anything about this problem to an in-house mechanism. In other words, the publishers themselves will have to require those who submit work based on quantitative data sets to make those sets available to other investigators. This can be done in one of two ways. The publisher might require those who submit work for consideration to agree to make their data available to others as a condition for final acceptance for publication. This is what the *Journal of Human Resources*, for example, does.¹ The alternative to this sort of arrangement would be a requirement that contributors who use data sets deposit a clean copy of the set with the publishers, providing written authorization to the publishers to make the data available to other researchers once publication has occurred. Obviously, this option is less attractive to publishers for much the same reasons that archival repositories and granting agencies are little interested in acting as clearing-houses. In some instances, however, this procedure might be necessary.

In dealing to this point with the questions of *who* is to decide and *how*,

1 The “*JHR* Policy on Replication and Data Availability” has as its second point: “Authors of accepted manuscripts will be asked to make their data available to other analysts from a date six months after *JHR* publication for a period of three years thereafter. Exemption from this requirement may be requested.” *JHR*, vol. 27, no. 3 (Summer 1992), p. 380.

I have not considered the question that is logically prior to all others. *Should* there be any mechanism at all to force gatherers of data to make their research materials available to others? The answer to this most important question, I think, is “Yes”, for at least three reasons. First, as the example I described at the outset of my remarks indicates, the absence of a mechanism to adjudicate disputes over control of and access to data is the source of delay, inconvenience, editorial anxiety, and – perhaps most important – possibly inequitable treatment of some investigators. Simple considerations of efficiency and equity argue strongly in favour of the creation of such a mechanism.

Second, there is the argument based on the public interest. If taxpayers have paid good money to support the collection of a data set held by someone who will not share it, they should not have to pay a second and a third time so that other scholars in the field can assemble data sets of their own with which to verify the conclusions of the hoarder. Funds for research are too scarce to contemplate any such procedures.

Third, and most important, the standards of historical investigation that we all claim to respect require that research results be tested. We cannot claim to be bona fide scholars if the results of our investigations are not subject to verifiability. If I publish work on Canadian government policy towards Aboriginal peoples, other researchers must be able to check the sources I cite in support of my conclusions. (Here too, however, there are growing problems. Work that relies on oral history, some of which has been collected after promises of anonymity have been given to the sources, is not always totally subject to verifiability of sources.) In the same way, researchers who argue for a decline in prices or living standards or productivity on the basis of sources that include quantitative data must make it possible for others to verify their analyses by providing full citations and, in the case of unique data sets, by making copies of those sets available. For historians of Canada to settle for anything less than such a procedure would be to exile themselves beyond the pale of legitimate scholarship.

“Whose data?” “Who decides?” These are important, pressing questions for historians in this country in the 1990s. There are valid reasons to be concerned about the lack of an accepted mechanism for deposit of and access to sets of quantitative data on which published results are based. Of the institutions or agencies that might act as such a mechanism, probably only publishers have the necessary power, or leverage, or clout to do so effectively. Everyone interested in healthy scholarship should support a move in the direction of having them do so.