The effects of electronic publication on social science researchers in Canada

Kirsti Nilsen
Faculty of Information Studies
University of Toronto
nilsen@fis.utoronto.ca

This paper considers the effects of electronic publishing on social science researchers, specifically as it relates to statistical data dissemination. Data of interest to social scientists are increasingly available only in electronic formats and, in Canada at least, at inflated prices. Canadian federal government cost-recovery and revenue generation initiatives of the mid-1980s encouraged government agencies to treat information as a marketable commodity. Statistics Canada responded to the initiatives with sharply higher pricing for its electronic and paper products. At the same time, much of the data formerly available in printed format became accessible only through electronic means. The effects of this policy on the use of Statistics Canada's data products by Canadian social science researchers is considered here, on the basis of research that examined the sources of statistical data used by authors of articles in Canadian social science research journals. Bibliometric analysis of citations, tables, and text of articles provides objective evidence of statistical data sources used and is complemented by a survey of authors.

Introduction

Electronic publishing is not limited to digitized texts and journals, but has been widely used by governments in the dissemination of statistical data. This paper considers the effects of this form of electronic publishing on social science researchers. It is particularly concerned with the extent to which these researchers use statistical data and any changes in use that may have resulted from the increasing dissemination of data in electronic formats and increasingly high prices charged for data. The paper begins by examining the literature on the use of electronic technology for dissemination of government information and statistical data and the implications of that use. Then Statistics Canada's use of this technology and philosophy of dissemination is considered. What follows is a discussion of findings of bibliometric and survey research that examined the use of statistics sources in five social science disciplines: economics, education, geography, political science, and sociology. The discussion focuses on the findings regarding electronic formats and pricing effects and the effects of the availability of funding. Data on individual disciplines are not provided.

Electronic publication of government information

The issues surrounding the application of information technology to government information and its effect on the government's provision of information are clearly complex. In a recent review of the literature, Hernon and McClure provide a long list of evolving issues related to the delivery of electronic information and services (Hernon and McClure 1993, 91–93). The application of technology to government information is by no means a negative process. This application has served to advance technology as governments have underwritten key technological innovations (Starr and Corson 1987, 434). The use of electronic technology has been beneficial to the acquisition, processing, and analysis of government information and has resulted in declining costs for the processing of large amounts of data, thus fostering a burgeoning information industry (Wallman 1988, 1). At the same time it has presented new means of dissemination, and it is this aspect which raises policy issues.

Hernon and McClure note that while there has been heavy government commitment in the United States to the use of technology for collecting, organizing and processing information, the technologies do not guarantee access to information. Further, they wrote that there had been "little formal attempt to assess the *specific impacts* of the increased Federal dependence on information technologies for the provision of government information to specific markets" (Hernon and McClure 1988, 35–36). In her dissertation research, Regan found that the net effect of computerized information systems in government is that they increase bureaucratic resources, reduce costs of processing information, and increase the value of information (Regan 1981, 28). One can conclude, then, that information technology is widely applied to government information because it provides real benefits in terms of information processing and because it enhances bureaucracy. Its external effects include an expanded information industry that involves itself as a stakeholder in the policymaking process. Public access to government information may be limited by technological barriers and by bureaucratic decisions regarding dissemination of government information stored in electronic formats.

Electronic publication of data

Government statistical data have been processed by computers since the 1940s, when the U.S. Census Bureau was involved in the development of the earliest electronic digital computers, ENIAC and UNIVAC I (Duncan 1987, 396). In Canada, Statistics Canada (then the Dominion Bureau of Statistics) saw the need for automated processing of data after the war and, by the time of the 1961 Census, was "well launched into the electronic and computer era" (Statistics Canada, *Annual Report* 1984/85, 3). Governments have processed data

electronically since these early days, but have begun to "publish" them electronically only in the last decade or so. Publication of data in electronic form has both positive and negative effects. On the one hand, data quality is significantly improved, and a very wide of range of data can be made available either as raw data to be analysed by individual researchers or in aggregated and summary form for those who prefer not to deal with the raw figures. On the other hand, negative effects arise from limitations on access that are implicit in the use of technology to disseminate data. Those who prefer to use paper products may find that certain data are no longer published in that format. Individual users may not have the technological or financial resources to access the electronic data, particularly if they are commercialized or privatized. There may also be bureaucratic or legal restrictions on data access that impede the progress of research.

Of all government information, statistical data are of value to perhaps the widest range of users including other levels of government, the private sector, nonprofit organizations, researchers, students, and the public at large, in addition to internal users. In the United States, statistical data encouraged the growth of the statistical services industry, which, as Starr and Corson (1987, 416) note, "probably would not exist if it were not for the long history of public investment in information technology and statistics." In Canada, because of Crown copyright and other restrictions, the industry has not been encouraged to the same extent. In the mid- to late-1980s, however, the information industry began lobbying government for access to government data. The Interdepartmental Working Group on Database Industry Support was established in the Department of Communications in 1988 to promote arrangements for private-sector dissemination of government-owned data (Morton 1995, 159–61; Morton and Zink 1991, 328; Nilsen, 1993, 205–8).

In the United States, information resources produced by federal statistical agencies are of substantial commercial value. Duncan wrote that these agencies generally followed the practice of charging only "costs of reproduction" to firms in the information industry. He noted that this policy raised the question of the cost-recovery basis, if any, by which the federal agency should determine charges against the private sector. Critics of the policy "argue[d] that more of the commercial value should be returned to the government through user fees that would recover more of the original program costs and better reflect the market value of products or service" (Duncan 1987, 405–6).

Wallman wrote of an expanded interest in user fees for statistical data in the late 1980s. She noted that the economic argument for such fees for government statistical data was made to the U.S. President's Economic Policy Council in 1987:

> charging fees for the use of federal statistics would capture some of the advantages of a market: if users had to pay for these products, they would have an incentive to be more discerning about the quality of data purchased. Producers, in turn would have an incentive to improve their products to meet users' requirements. The prices of statistical data would reflect their market value. (Report of the Working Group 1987, in Wallman 1988,11)

With governments willing to consider market options for their products and services, the debate over commercialization or privatization, technology implications, and public access to statistical data has intensified in the 1980s and 1990s. The larger issue here is the question of why statistical data are collected by governments. Data are collected to fulfil internal needs and to gather and disseminate information required by law or for public purposes. Starr and Corson (1987, 446) wrote that the most serious risk of privatization of government data was diminished public access: "The price of private information services puts them beyond the reach of most community organizations, inquisitive citizens or academic researchers."

The concerns regarding limits on public access and its effects on research are eloquently expressed by Starr and Corson, (1987, 447):

> The privatizing of statistical information poses some specific problems for the future of the social sciences and intellectual life. Without the capacity to make use of the new information resources in private hands, the universities and other nonprofit research centers may be left as intellectual backwaters. Such a development would have wide implications. Our society maintains universities and research institutions because they create new knowledge and help introduce it into education. The rules of the academy favor open conduct of research; in this sense, all such knowledge is public. Insofar as the public domain is restricted in the sphere of knowledge, research in the university suffers, and insofar as the research in the university suffers, so does public knowledge. The more the national information base passes into the proprietary sphere, the more constricted will be our resources for public life and social inquiry.

In Canada, the information industry has not been given access to government data to the extent seen in the United States. However, Statistics Canada has aggressively marketed its data on a cost-recovery basis that has resulted in impressive price increases. At the same time, it has moved strongly into electronic dissemination of its data at the expense of paper publication.

Statistics Canada

Statistics Canada has a wide mandate that requires it to "collect, compile,

analyze, abstract and publish statistical information relating to the commercial, industrial financial, social, economic and general activities and condition of the people." It collects its data from individual Canadians, from businesses, and, through joint agreements, from other Canadian governments and other federal agencies. Individuals and businesses are legally required to collaborate with Statistics Canada in its data collection (Canada, 1985, Statistics Act, Ch.S-19, ss.3a, 31).

The agency has responsibility for the production of most of Canada's statistics, and is, in effect, in a monopoly position with respect to the data it collects and disseminates. However, it has traditionally seen its information as a "public good," which was to be widely available at minimal cost:

> By virtue of having underwritten through taxation the costs of Statistics Canada's operations and having collaborated in furnishing the raw material of its finished statistics, the bureau's users have a proprietary interest in such statistics. These are seen as a "public good" with a high threshold of free or nominally-priced accessibility. (Statistics Canada, *Annual Report* 1979/80, 1).

In a 1990 article, Douglas Newson, then Director of its Communication Division, wrote that the word "publish" in the mandate "has always been interpreted to include data in all media. The *raison d'être* of the organization thus includes data dissemination...[which] has always been viewed as a legitimate part of our activity" (75).

Nevertheless, since the late 1970s, when the "public good" view of its information was expressed, Statistics Canada has undertaken a number of initiatives that suggest that it is interpreting its mandate more narrowly. This new direction appears to be an outcome of years of budget cutbacks and an increasing appreciation of the monetary value of its information. By the late 1980s, an *Annual Report* described the ongoing challenge of "enhancing [the] product line...while operating in an environment of fiscal restraint and reductions in staff and resources." This involved "a corporate rethinking of Agency priorities and goals and a new more streamlined approach to production and outputs" (Statistics Canada, *Annual Report* 1988/89, 4–5). Such rethinking includes distribution of an ever-smaller range of "humanized" (i.e., more accessible, graphic, and interesting), but more expensive, print products to the general public and an ever-increasing profitable distribution of electronic products and customized outputs to experts and specialized users (Statistics Canada *Annual Report* 1988/89, 6). Dissemination of data in electronic formats was increasingly emphasized, particularly to "informed" users who could access electronic data directly (Statistics Canada *Annual Report* 1985/86, 5;

1986/87, 2, 17; 1987/88, 20). Electronic dissemination of data is seen as a means of revenue generation.

Budget restrictions

Statistics Canada began experiencing budget restrictions in the later 1970s. In the mid 1980s, the Conservative government introduced across-the-board restraint initiatives and a cost-recovery policy. Government publications were specifically cited as an area that "provides opportunities for the government to realize a larger flow of revenues from those individuals or groups who benefit most directly from these services" (Canada, 1985, Budget Papers, 29–30; Canada, 1986, Budget Speech, 7–9; Canada, 1986, Budget Papers, 5). The agency's annual budget was reduced by $2.5 million, and it was required to increase its revenues by $1.5 million and to implement program output reductions of $1 million (Statistics Canada, *Estimates* 1987/88, 20). It was expected to undergo reduction and cost-saving commitments of $26 million over the period from 1985/86 to 1990/91. Although there have been more recent budget cuts, it was this mid-1980s move to encourage cost-recovery and "business-like management practices" that has had the most profound effect on data dissemination. Statistics Canada was encouraged to see its information as a commodity and to market its products with revenue generation in mind. At the same time it began to look at a "more balanced role as both a direct distributor of data to the end user and as a provider of data to the private information industry for redistribution" (Newson 78).

Prices

Statistics Canada has responded to budget cuts over the years by cutting staff, merging and eliminating information products, and raising prices. Its 1984/85 *Annual Report*, written after the restraint and cost-recovery initiatives were announced in 1985, stated that it had:

> set prices for its data products to recover such costs as printing, mailing and handling, in the case of printed publications or corresponding charges for computer staff time when special tabulations are requested in other formats such as microfilm, microfiche or computer printouts... [The changes] do not reflect the costs of maintaining the system for compiling information and the methodology and analysis that make the data valuable in the hands of its users. These are maintained mostly as a public good. (7)

In 1985/86, the agency implemented the policy of user-pay and cost recovery for publications, electronic data dissemination, and special statistical products and services. It hoped to achieve its goal of cost-saving by (among other things)

improving "further the cost-revenue relationships of the agency's products and services" (Statistics Canada, *Estimates* 1986/87, 7).

It should be noted that the agency does provide some information free of charge:

> Statistics Canada has a longstanding policy to provide public access free of charge to general statistical data. This objective is achieved through a variety of means including: the Depository Library Program; provision of the Statistics Canada *Daily* to the news media to ensure general news coverage; provision of reference librarians in Statistics Canada regional offices and provincial/ territorial statistical offices; responding to over half a million requests for general information each year, primarily by telephone. (Newson, 76)

To imply, as Newson does, that these activities are sufficient to fulfil the dissemination requirements of its mandate, suggests that the agency is narrowly interpreting the Statistics Act.

Newson went on to state:

> However, in accordance with general government policy, users who require their own private copy of a Statistics Canada product/service are now required to pay the full costs of making that product or service available to them... [The agency] requires clients to pay for the costs of production and distribution of the products. Prices are set in such a manner that the full cost of dissemination is recovered—not necessarily for each product but for the entire range of output of the agency. This distinction between the public good and a private good is fundamental in our dissemination activity. (Newson, 76)

In 1990/91 the agency stated that it had achieved balance between publication costs and revenues: each totalled $7.4 million. This break-even position was attributed to "strong subscription sales; more aggressive marketing of quarterly, semi-annual and annual publications; better tailoring of products to user needs; and increased printing efficiencies" (Statistics Canada, *Annual Report* 1990/91, 25).

A comparison of prices over the years for individual products and for average prices shows that price increases have undoubtedly been a major influence in the achievement of these revenues. For example, the agency's flagship publication, *Canadian Economic Observer* (while still titled *Canadian Statistical Review)*, jumped from an annual subscription rate of $38.50 in 1984 to $200 in 1985. Since then the price has climbed to $220. An examination of prices of print products over the period 1982–1993 shows average increases of 208%. The maximum price for a print product in 1982 was $75; in 1993 it was $551, an increase of 1361%. One publication increased by 3005%.

Prices for electronic products have risen at an ever faster rate. In a report on electronic product pricing, Laine Ruus, now Data Librarian at the University of Toronto, noted that the cost of 1986 computer-readable census products had increased between 1,567% and 9,900% in three years (Ruus 1987, 1). These prices led Canadian research universities to form a joint purchasing consortium, which obtained special pricing arrangements for electronic census products as well as some other categories of data. However, microdata files were not included. Average prices for data in these microdata files can range from $500 to $2,000 and up for one year of data (Statistics Canada, *Catalogue*). Even though universities were able to get a 50% discount on data purchases, the cost was often beyond their means.

Laine Ruus has frequently raised the point that users should be charged less for electronic data than for the paper products. Though increases in prices of paper products are not desirable, her point is well taken. For raw data files, in particular, editorial costs are practically non-existent. The data have already been gathered and processed electronically; and where the agency is charging excessively high prices, one can argue that it is attempting to recover more than the costs of dissemination. Starr and Corson have noted that while "the costs of storing and manipulating data have been falling, the costs of collecting it in the first place have been growing" (1987, 423). Since this is undoubtedly true, one wonders if Statistics Canada is not somewhat disingenuous in its claims that prices are based only on dissemination costs.

More recently there have been encouraging signs that Statistics Canada recognizes the need to make its data more widely available at less cost to social science researchers. As of early 1996, some microdata files and other materials are available to university researchers as a result of the Data Liberation Initiative begun by the Social Science Federation of Canada and the Canadian Association of Research Libraries. By paying a single set price, universities will be able to provide researchers with electronic access to a wide range of data for non-commercial purposes.

Data restrictions

There are various ways that data restrictions can limit access to information. Legal restrictions set out in the Statistics Act present problems for some users of Statistics Canada data. These restrictions limit disclosure of data that can be related to a specific respondent. "Statistics Canada places the highest priority on ensuring that the rights of respondents are respected" with regard to confidentiality (Statistics Canada, *Annual Report* 1987/88, 2). Social science researchers comment that these restrictions are onerous and slow the progress of research. As one respondent to a survey by this researcher stated, "Confiden-

tiality restrictions impede the utility of data released because often the categories are too crude" (Nilsen 1996).

Limits on access caused by agency decisions are more subtle. At the same time it was increasing prices, Statistics Canada was cancelling print publications or merging them into more general broadly based products. In 1988 it wrote that a review of its publications recognized that "most users of publications are interested in general information, rather than in highly specialized data." This attitude was expected to result in a smaller publications program, with more publications addressed to non-specialists. "Detailed information will, of course, continue to be available on a cost-recovered basis" (Statistics Canada, *Annual Report* 1987/88, 13). In practice, where formerly one could subscribe inexpensively to a narrowly focused statistical product, now a group of related products is published in larger, easy-to-read, more attractive, and more expensive "compendia," designed for general interest. Those who seek the detail formerly available in the narrowly focused product must request and pay for "specialized" information in electronic form on a cost-recovery basis.

Research results

The research described here sought to evaluate the effects of Statistics Canada policies on social science researchers in Canada. This paper describes the findings with respect to the electronic publication of data. Two research approaches were used: first, a bibliometric analysis of a sample of articles published in Canadian social science research journals; and second, a survey of authors of articles in the sample. In the discussion that follows, all percentages have been rounded.

Bibliometric analysis

The population for the bibliometric analysis is a sample of 360 articles from 21 Canadian peer-reviewed journals published over the period 1982–1993 in the disciplines of economics, education, geography, political science, and sociology. These disciplines were identified through an examination of the literature on use of materials by social scientists and in specific disciplines, use of government publications by social scientists, and citation studies. Research reviewed included the ambitious 1971 *Investigation into Information Requirements of the Social Sciences* (INFROSS) to which 1,089 social science researchers in the United Kingdom had responded (Line 1971), the 1979 study *Design of Information Systems in the Social Sciences* (DISISS), Fletcher (1972), Jarvi (1976), Hernon (1979), Fitzgibbons (1980), Brittain (1982), Hernon and Purcell (1982), Hernon and Shepherd (1983), Meyriat (1984), Matsuda and Suoh (1986), and Slater (1989).

An examination of text, tables, and citations in the articles in the sample provided objective evidence of use of statistical data sources and of any changes over time in that use. The discussion here examines the use of statistics in the 207 articles that had a Canadian focus or setting and that used published statistics. Published statistics, in this context, include data collected and made available by a source other than the researcher him/herself. The data may be analysed or abstracted by the data producer, or the researcher may have access to the raw figures.

Use of statistics sources

Table 1 illustrates the variety of statistics sources used in articles written with a Canadian focus or setting. By far the most heavily used source of data was nongovernmental sources, which included statistical products of universities, businesses polling organizations, and associations as well as trade and university press monograph and journal publishers. Canadian nongovernmental sources were used in 117 articles (57%) and American nongovernmental sources in 67 articles (32%).

That there was no statistically significant variation over the period 1982 to 1993 in the number of articles using the various statistics sources suggests that social science researchers were consistent in the sources they went to when seeking published statistics and that changes in formats and prices for Statistics Canada data did not lead them to change sources.

Use of Statistics Canada materials by format

Of the 85 articles with a Canadian focus or setting that used Statistics Canada as a data source, 71 (84%) used paper products and 29 (34%) used computer-readable products. (Some used both formats.) This is a ratio of 2.5 to 1.

Variation on a year-to-year basis over the period 1982 to 1993 in the number of articles using these products was not statistically significant. However, if one compares the first half of the period (i.e., 1982–87) with the second half (1988–93), there is variation that suggests that use of paper products is declining and use of computer-readable products is growing (see Table 2). The decline in use of paper products is statistically significant.

The percentage of articles using computer-readable formats as a data source was highest in the last two years studied. These formats were used in 20% of the articles in 1992 and in 36% of the articles in 1993. Overall, these data suggest a growing use of computer-readable formats.

Variation by funding

Of the 85 articles that used Statistics Canada as a source of data, only 26 (30.6%)

Table 1. Statistics sources used (N=207)

Source	No.	%
Statistics Canada	85	41
Other Canadian federal government sources	84	41
Provincial government	52	35
Municipal/Regional government	18	9
Foreign government (all)	24	12
United States government	21	10
Intergovernmental organizations	11	5
Nongovernment sources	147	71

Table 2. Number of articles using Statistics Canada products by format (N=85)*

	Paper ($\chi 2$ =.045)	Computer readable ($\chi 2$ = .290)
1982–1987	36	11
(n=39)	(92%)	(28%)
1987–1988	35	18
(n=46)	(76%)	(39%)

*85 articles which used Statistics Canada as a data source.

indicated in the article that the research had been funded. While there could have been funding sources that were not noted in the article, the assumption is that funding sources are routinely thanked.

Of the 71 articles that used paper products, 21 (29.61%) were funded, and of the 29 articles that used computer-readable formats, 13 (44.8%) were funded. It appears that being funded may incline researchers to use computer-readable data. However, there was no statistically significant relationship between funding and format used.

Number of paper and computer-readable products used

The above analyses have looked at use of paper and computer-readable formats in articles that used Statistics Canada as a data source. The number of these products used in each article is of interest. While the percentage of authors using computer-readable products has grown, to what extent are more computer-readable products being used?

The 85 articles that used Statistics Canada as a data source used a total of 236 paper products and 50 computer-readable products. However, analysis showed that variation over time in the number of paper and computer-readable products used was *not* statistically significant. While computer-readable products were used in a greater number of articles following 1987 than earlier, the number of products used per article has not grown.

Survey results

A survey was mailed in the fall of 1995 to 163 Canadian academics who had been identified in the bibliometric research as having written articles with a Canadian focus or setting that had used published statistics; 97 completed surveys were returned, a response rate of 60%. This paper provides an analysis of the responses with respect to use of electronic data and pricing of those data.

Extent of use of statistics

Responses to the question "To what extent do you use statistical data (i.e. numeric information) in your research or writing?" are provided in Table 3. We can see that 71 (74%) of the respondents used statistics more than 50% of the time in their research or writing. The remaining 25 (26%) used statistics only occasionally or seldom.

Statistics sources used

Asked to "indicate the sources of the statistical data which you have used or currently use," respondents answered as shown in Table 4. In comparing the data to the bibliometric data in Table 1 above, one sees that Statistics Canada was used in 41% of the articles, while 86% of survey respondents indicated that they had used Statistics Canada as a data source at some time. On the other hand, nongovernmental sources were used in 71% of the articles, but only 40% of survey respondents indicated that they had used them. Use of Statistics Canada as a data source may be greater than the bibliometric research indicated, for this research looked at specific articles, whereas the survey asked for "lifetime" use of data sources. In Table 11 below, which indicates frequency of use of Statistics Canada, the findings more closely correlate with the bibliometric findings. The use of nongovernmental sources is undoubtedly greater than the survey respondents suggested. The bibliometric analysis closely looked at every data source used, while survey respondents may have been thinking in terms of major uses of data sources as opposed to all uses of statistics.

Table 3. Extent of use of statistics (N=96)

	No.	Percentage*
Seldom [up to 20% of the time]	12	12.5
Occasionally [21-50% of the time]	13	13.5
Often [51-80% of the time]	26	27.1
Almost always [81-100% of the time]	45	46.9

*Percentage of 96 who answered the question.

Table 4. Statistical sources used (N=97)

Statistical Source	No.	Percentage
Statistics Canada	83	86
Other Canadian federal government sources	56	58
Provincial governments	54	56
Municipal or regional governments	18	19
French government	1	1
United Kingdom government	7	7
United States federal government	36	37
United States state or local government	5	5
Intergovernmental organizations	27	28
Non-government sources	39	40
Collect my own data	60	62
Other*	9	9

*One respondent wrote "other Western European governments" and two respondents used government statistics of Cuba and Mali. Five respondents indicated use of other foreign sources of data (i.e. Germany, Japan, Denmark, and "developing countries." It is not clear whether these were government or nongovernmental sources. Three respondents did not identify the other sources.

Formats

To the question "Have you ever use any computer-readable data for your research?" 79 (81%) respondents answered Yes and 18 (19%) No. It would appear that a relatively small number have never used computer-readable data. Responses to "In what formats have you normally obtained the data that you use?" are shown in Table 5. A large percentage are still obtaining data from print sources, but computer-readable files are also an important source. Some respondents indicated that they normally used all of the formats named.

Respondents were asked to identify the single format they would most prefer to use. As this question attempted to identify preferences in formats of published statistics, no option was available for those who most prefer to collect their own data. Nor were microforms an option. Of the 77 who provided usable responses, almost 65% indicated they would prefer to use computer-readable files, but many (almost 30%) still prefer to use print (paper) formats.

In comparing Table 6 to Table 5 above, note that 76% normally use paper, but only 29% would most prefer to use paper. That the percentages of those who normally use computer-readable files (61%) and those who most prefer them (65%) are close suggests that more will be using them in the future. While 37% normally use special tabulations, only 1% most prefer them. The difference here may relate to price. Special tabulations from Statistics Canada are expensive, but may be a necessary source for researchers, for these data are not available elsewhere.

How data are normally acquired

Responses as to how the researchers normally acquired the data they used are shown in Table 7. Asked to rank their first three preferences of the various means of acquiring data, they responded as shown in Table 8. Note that where 47% in Table 7 use a library to acquire paper copies, for only 20% is that a "top three" preference. Many more would prefer to purchase computer-readable files than actually do. And more would prefer to use a data library. But fewer would prefer to collect their own data than actually do.

Funds for purchase of computer-readable data

Asked to identify the source of the funds used to purchase computer-readable data, if they had ever done so, 70 responded. The assumption is that they would not have answered this question if they had never purchased computer-readable data. However, according to Table 7 above, only 50 normally acquire data in this way. Thus, for 20 respondents the purchase of computer-readable data appears to be an exception rather than the norm. (See Table 9.) The paramount importance of grant money in the purchase of computer-readable data is evident

Table 5. Formats in which data are normally obtained (n=97)

Formats	No.	Percentage*
Print (i.e. paper)	74	76
Microform (microfiche, microfilm)	13	13
Computer-readable files**	59	61
Special tabulations***	36	37
Collect my own data	60	62

*Numbers exceed 100% because respondents could use more than one format.
** Computer-readable files were defined as "any 'off-the-shelf' computer files created or public dissemination or for limited dissemination with business or commerce, etc.
***Special tabulations are created in response to specific request for a specific purpose, whether in print or computer-readable form.

Table 6. Most preferred format (N=77)

	No.	Percentage*
Print (i.e. paper)	23	29
Microform (microfiche, microfilm)	0	0
Computer-readable files	50	65
Special tabulations	1	1
Other**	3	4

*Percentage of 77 usable answers.
**Two of those who chose "other" wrote that it depended on the particular project, and the third respondent indicated preference for transcripts and case studies.

Table 7. How data are normally acquired (N=97)

Means of acquiring data	No.	Percentage
Purchase paper copies	35	36
Purchase microform copies	9	9
Purchase computer-readable files	50	52
Purchase special tabulations	30	31
Use a library for paper or microform	46	47
Use a data library	37	38
Use a departmental collection for paper copies	10	10
Use a departmental collection for computer-readable data	15	16
Use the Internet	19	20
Collect my own data	56	58
Other	2	2

Table 8. Ranked preferences for acquiring data (N= 79)

Means of acquiring data	First choice	One of top three choices
Purchase paper copies	10%	27%
Purchase computer-readable files	23%	68%
Use a library for paper copies	10%	20%
Use a data library	20%	52%
Use the Internet	3%	23%
Collect own data	23%	40%

Table 9. Source of funds for purchase of computer-readable data (N=70)

Source of funds	No.	Percentage*
Personal funds	7	10
Funded grant money	58	83
Departmental/ institutional money	25	36
Other	3	4

*Total exceeds 100% because more than one source of funds could be identified.

Table 10. Extent of use of Statistics Canada data in last ten years (N=94)

Extent of use	No.	Percentage
Seldom (up to 20% of the time)	28	30
Occasionally (21-50% of the time)	20	21
Often (51-80% of the time)	28	30
Almost Always (81-100% of the time)	11	12
Did not use	7	8

in these responses. If grants are not available or do not cover the cost of data in this format, researchers will be hard put to purchase them.

Use of Statistics Canada data

In the analysis that follows, there is no distinction between formats of Statistics Canada data. Respondents may be using paper, computer-readable, or both formats.

Extent of use

Responses to indicate the extent of use of Statistics Canada data in research in the last ten years seem to support the bibliometric findings (see Table 10). There were 39 (42%) who used Statistics Canada data "often" or "almost always" and 55 (58%) who used them "occasionally" or "seldom," or did not use them. The percentages are very similar to the documented use of Statistics Canada in the articles examined for the bibliometric research. Table 1 shows that 41% of the articles used Statistics Canada as a data source and 59% of the articles did not.

In order to compare use in the last ten years with use before 1985 (when prices jumped), respondents were asked to recall extent of use of Statistics Canada data at that time. The figures indicate that 32 (35% of those responding) used Statistics Canada data "often" or "almost always" and that 59 (65%) used these data "seldom" or "occasionally" or not at all. When compared with the figures in Table 10, these responses would seem to suggest that use of Statistics Canada data has increased. However, the number who responded that they did not use these data before 1985 may be unreliable, for 20 respondents only began to do research in the 1980s and eight of these began in 1984 or later. In order to best compare the two periods, the number who did not use Statistics Canada data were excluded, and the percentages based only on those who did use them. Table 11 compares the two periods. This table shows that the percentage of respondents who used Statistics Canada data "seldom" or "occasionally" did not change over the two periods. Possibly there is a consistent 55% of researchers for whom Statistics Canada data are not very useful. Those who used these data "often" or "almost always" remained at 45%, which suggests that there has been no real growth here: while the percentage using the data "often" increased slightly, the percentage who used them "almost always" decreased. The percentage change is not great, but one wonders if price has had something to do with the decrease in those who use Statistics Canada data "almost always" (i.e., more than 80% of the time).

Funding vs. use in the last ten years

In order to determine whether the ability to get funding can be related to use of

Statistics Canada data, a comparison was made between the extent of use of these data and funding of recent research (Table 12). Of those who had been funded in their recent research, only 40% had used Statistics Canada data in the last ten years. Thus, ability to be funded is not necessarily related to the use of Statistics Canada data in general.

Responses to a question that asked for reasons for any change in use of Statistics Canada data over the years are shown in Table 13. Almost 50% indicated that their research agenda had changed. Reasons for such change were not requested and not provided. It is possible that availability of data could have led to research agenda change, in either a positive or negative sense. Six respondents wrote that they are using more data than ever, because the quality and quantity of the data has improved. Two respondents indicated that they were using more data "despite the price." However, the fact that over 20% of those whose use of Statistics Canada data changed cited the cost should be of concern to the agency. Driving away customers is not good marketing strategy.

Those who "seldom" (i.e., less than 21% of the time) use Statistics Canada data were asked to indicate why. Although 41 responded to this question, only 28 had earlier indicated that they seldom use Statistics Canada data (see Table 10). As Table 14 shows, the 25 who responded that "there are no data applicable to my research" (61% of those who "seldom" use Statistics Canada data) represent 32% of the total sample. In Table 10 above, there were 28 who

Table 11. Comparison of extent of use of Statistics Canada data

Extent of use*	Before 1985 (n=71)	Last ten years (n=87)
Seldom	22 (31%)	28 (32%)
Occasionally	16 (23%)	20 (23%)
Often	19 (27%)	28 (31%)
Almost always	13 (18%)	11 (13%)

*See Table 10 above for definition of terms.

Table 12. Use of Statistics Canada data by those who have been funded (N=82)*

Extent of use**	No.	Percentage
Seldom	23	28
Occasionally	19	23
Often	22	27
Almost always	11	13
Did not use	4	5

*Number whose recent research has been funded.
**See Table 11 above for definition of terms.

Table 13. Reasons for any change in use of Statistics Canada data over the years doing research (N=84)

Reasons	No.	Percentage
No change in use of Statistics Canada data	30	36
My research agenda changed	41	49
Less information is available now than formerly	4	5
Other sources of data more useful to my research	11	13
Paper copies of data I want are no longer available	5	6
Statistics Canada data are too costly	19	23
Computer-readable data are too costly	5	6
Computer-readable data are not easily accessible	1	1
Funding is insufficient to cover cost of Statistics Canada data I need	9	11
Other **	9	11

**Six wrote they are using more data than before.

Table 14. Reasons why Statistics Canada data are seldom used (N=41)

Reasons	No.	Percentage
There are no data applicable to my research	25	61
The data which I want are too expensive	7	17
The data which I want are not available in paper	1	2
The data which I want are aggregated at a level inappropriate to my research	4	10
The data which I want are available only in unaggregated form	1	2
The data which I want are available only through special tabulations	5	12
I prefer other Canadian sources of data	5	12
I prefer other non-Canadian sources of data	1	1
Other*	4	10

*One wrote "lack of funding for secondary analysis," others doing theoretical research.

"seldom" use Statistics Canada data. Thus it appears that most of those who seldom use its data do not find them applicable to their research.

Price changes

Asked to what extent they were affected by the price changes of recent years for Statistics Canada data, they responded as shown in Table 15. In all, 60 (63%) responded that the question was not applicable to them, or that they were not aware of prices changes, or they were unaffected by them. Presumably there is little overlap in this group, and one can conclude that these respondents either make little use of Statistics Canada data or have sufficient funding to be able

Table 15. Extent affected by price changes (N=96)

Effects of price changes	No.	Percentage
Not aware of changes in prices	19	20
Aware of price changes but unaffected by them	29	30
Price changes have affected my allocation of available funds	20	21
My research has been adversely affected by price changes	20	21
My research agenda has changed because of price changes	7	7
My data sources have changed because of price changes	9	9
Other*	4	5
Not applicable	12	13

*One respondent wrote, "must spend more time in grant applications to get money to buy data." Another wrote that s/he must make "more selective, less exploratory purchases." The third respondent (a demographer) cited the Inter-University Consortium as compensating for price increases. The fourth respondent argued that his/her "research has been positively affected by price changes..data are relatively much cheaper, and more readily available."

Table 16. If Statistics Canada data use has changed because of price changes: what has happened? (N=36)

	No.	Percentage*
Used less Statistics Canada data	22	61
Used other Canadian sources for statistical data	9	25
Used non-Canadian sources for statistical data	8	22
Collected data myself	8	22
Other**	7	19

*Percentage of 36 who answered the question.

**One respondent wrote that s/he "abandoned the subject" and another, "I have spent less for research assistants." A third respondents wrote that s/he had "not conducted the preferred research" A fourth wrote, "I have been more efficient in my choices." A fifth wrote the s/he "used up my research time by contacting public libraries." Two respondents stated that they relied on data provided through the Consortium. One respondent wrote that on the contrary s/he used more Statistics Canada data. This was the same respondent who suggested that prices were relatively cheaper because the data quality was much improved.

Table 17. Comparison of percentage use of any computer-readable data with percentage using Statistics Canada computer-readable data (N=97)

	Yes	No
Ever used any computer-readable data?	79 (81%)	18 (19%)
Ever used Statistics Canada computer-readable data?	57 (59%)	40 (41%)

Table 18. Funding vs. use of Statistics Canada computer-readable data (N=96)

Recent research funded?	Yes	No
Used Statistics Canada computer-readable data at some time*	52 (64%)	5 (36%)
Never used Statistics Canada computer-readable data	30 (37%)	9 (64%)

*Use was not necessarily for the recent funded research.

to purchase necessary data. Of 36 respondents (35%) who indicated effects of price increases, several noted more than one effect. Of particular interest is the 20% who suggested that price changes had affected their allocation of available funds. From several comments, it appears that some researchers no longer can afford to hire graduate students to work on the data.

Table 16 shows responses to a questions about what has happened if there has been a change in use of Statistics Canada data as a result of price changes.

Format changes

Respondents were asked if they had ever used computer-readable data and if they had ever used Statistics Canada computer-readable data. Table 17 illustrates the use made of Statistics Canada computer-readable data by those who have ever used any computer-readable data. With more than 80% of respondents indicating that they had used computer-readable data, the percentage who had not used Statistics Canada's data is surprising. The reason could be price or it could be the inappropriateness of the data to some researchers. Given that Statistics Canada is the major supplier of computer-readable data in Canada, one would expect a higher number of Canadian social science researchers to use them.

The importance of funding was examined by correlating the use of Statistics Canada computer-readable data with responses to a question as to whether the research in which respondents had been involved in the last three years was funded. It was found that those who had been recently funded were more likely to have used these data than were those who had not (Table 18).

While these findings suggest a correlation between availability of funding and the use of Statistics Canada data, they are *not* significant. The fact that more than a third of those who had been funded had never used this data source suggests that funding is not the only prerequisite for use.

In order to determine whether more experienced researchers were more likely to use Statistics Canada computer-readable data, use was correlated with the year researchers began to do research. The results are provided in Table 19. The data show that those who began to do research since 1980 are much more likely to have used Statistics Canada data than those who started earlier. If one can correlate "year began to do research" with age, one can conclude that younger researchers are more likely to have used these data than are older ones.

Researchers were asked the extent to which they were affected by format changes (Table 20). If one assumes no overlap in the responses "unaware of format changes," "aware but unaffected," and "not applicable," one can conclude that in total 67 respondents (70%) were unaffected by format changes.

Table 19. Date began to do research vs. use of Statistics Canada computer-readable data (N=96)

Year began to do research	Before 1970 (n=30)	1970–79 (n=47)	1980–89 (n=20)
Used Statistics Canada computer-readable data at some time	14 (47%)	26 (55%)	17 (85%)
Never used Statistics Canada computer-readable data	16 (53%)	21 (45%)	3 (15%)

$\chi^2 = .007$

Table 20. Extent affected by format changes (N=94)

Effects of format changes	No.	Percentage
Not aware of any format changes	16	17
Aware of format changes but unaffected by them	35	37
My research has been positively affected by format changes	18	19
My research has been adversely affected by format changes	6	6
My research agenda has changed because of format changes	2	2
My data sources have changed because of format changes	2	2
Other*	3	3
Not applicable	16	17

*Two described negative effects, one did not comment.

Table 21. If Statistics Canada data use has changed because of format changes: what has happened? (N=35)

	No.	Percentage*
Used a greater variety of Statistics Canada print sources	5	14
Used more Statistics Canada computer-readable data	20	57
Used more Statistics Canada special tabulations	7	20
Used other Canadiann (non-StatsCan) print sources	8	23
Used other Cdn (non-StatsCan) computer-readable sources	2	6
Used other non-Canadian print sources	2	6
Used other non-Canadian computer-readable sources	1	3
Collected data myself	2	6
Other**	2	6

*Percentage of 35 who answered the question and found it applicable to them.
**One indicated that s/he used Statistics Canada data but accessed them less frequently.

There is some overlap in the responses of the remaining 29; of these only 6% suggest clear adverse effects of format changes.

Asked what has happened if their use of Statistics Canada data had changed because of format changes, most indicated an increasing use of computer-readable data (Table 21). Among those who indicated an increased use of the

agency's computer-readable data, one wrote that there is "all sorts of new stuff available—*never* in print!" and another cited availability of microdata. It appears that the response to format changes has been largely positive.

Opinions

Respondents were asked their extent of agreement or disagreement with two statements related to Statistics Canada computer-readable data and to prices. The responses are provided in Table 22. With 88% of those who have an opinion responding that they agree or strongly agree that Statistics Canada computer-readable data are helpful to their research, one can conclude that in general respondents are pleased with the trend to make more and more data available in computer-readable form. However, the fact that 27% indicated that they have no opinion suggests that there is still more than one-fourth of researchers who have no need for these data or who prefer not to use them for other reasons.

Price is obviously an issue (Table 23), with 70% of those who have an opinion disagreeing with the statement that Statistics Canada prices are reasonable. Presumably a large percentage of those who find the data helpful to their research are unhappy with the prices charged for them.

Table 22. Statistics Canada computer-readable data are helpful to my research

	No.	Percentage of those responding (n=89)	Percentage of those with an opinion (n=65)
Strongly agree	28	32	43
Agree	29	33	45
Disagree	3	3	5
Strongly disagree	5	6	8
No opinion	24	27	-

Table 23. The price of Statistics Canada data is reasonable

	No.	Percentage of those responding (n=90)	Percentage of those with an opinion (n=58)
Strongly agree	0	0	0
Agree	17	19	29
Disagree	20	22	34
Strongly disagree	21	23	36
No opinion	32	36	-

Conclusion

The research results presented in this paper are selected from a mass of data that will undergo further analysis. The data examined here suggest that the price increases for Statistics Canada products have not had an effect on the data sources used or the number of data products used by social scientists. The objective bibliometric research found that Statistics Canada data had been used in only 41% of the articles written with a Canadian focus or setting that used statistics. The bibliometric research found no changes over time in use of the agency as a data source. Survey respondents indicated greater use of Statistics Canada, but the number who use it "almost always" (i.e., more than 80% of the time) has declined since the mid-1980s. Other effects of price increases may be more indirect. For example, some researchers have ceased hiring graduate students to assist with data analysis. If fewer graduate students are hired to work on these data, they in their turn may use other sources in the future.

It is clear that researchers are unhappy with prices charged for Statistics Canada data.

Increased availability of data in electronic formats appears to be well received, and the use of paper products is declining. How much this decline results from the fact that fewer paper products are available has not been analysed. Availability of funding appears to be related to the use of computer-readable data in general, but is not the only prerequisite for the use of Statistics Canada data.

References

Brittain, Michael (ed.). 1982. *Social sciences: the supply of and demand for documentation and data*. A report to UNESCO. London: Rossendale.

Canada. 1985. Budget Papers. Tabled in the House of Commons by Michael Wilson, Minister of Finance. Ottawa: Dept. of Finance. May 23.

Canada. 1985. Statistics Act. Ch. S-19. Revised Statutes of Canada. Ottawa: Queen's Printer.

Canada. 1986. Budget Papers. Tabled in the House of Commons by Michael Wilson, Minister of Finance. Ottawa: Dept. of Finance. February 26.

Canada. 1986. Budget Speech. By Michael Wilson, Minister of Finance. Ottawa: Dept. of Finance. February 26.

Design of information systems in the social sciences. 1979. [DISISS]. The Structure of Social Science Literature as Shown by Citations. Research Report A no.3. by Maurice Line. Bath, U.K.: Bath University Library.

Duncan, Joseph W. 1987. Technology, costs, and the new economics of Statistics. In *The politics of numbers*, ed. William Alonso and Paul Starr, 395–413. New York: Russell Sage Foundation.

Fitzgibbons, Shirley A. 1980. Citation analysis in the social sciences. In *Collection development in libraries: a treatise.* Greenwich, CT: JAI Press

Fletcher, John. 1972. A review of the literature of economics. *Journal of Documentation* 28 (4): 283–95.

Hernon, Peter. 1979. *Use of government publications by social scientists.* Norwood, NJ: Ablex.

Hernon, Peter, and Charles R. McClure. 1993. Electronic U.S. government information: policy issues and directions. In *Annual review of information science and technology,* vol. 28, ed. Martha E. Williams, 45–110. Medford, NJ: Learned Information for ASIS.

Hernon, Peter, and Charles R. McClure. 1988. *Public access to government information: issues, trends, and strategies.* 2nd ed. Norwood, NJ: Ablex.

Hernon, Peter, and Gary R. Purcell. 1982. Document use patterns of academic economists. Chapter 5 in *Developing collections of U.S. government publications.* Greenwich, CT: JAI Press.

Hernon, Peter, and Clayton A. Shepherd. 1983. Government publications represented in the Social Sciences Citation Index: an exploratory study. *Government Publications Review* 10: 227–44.

[INFROSS]. *Investigation into information requirements of the social sciences.* 1971. Research Report no.1. Information Requirements of Researchers in the Social Sciences. Maurice B. Line, Project Head. Bath, U.K.: Bath University Library. 2 volumes.

Jarvi, Edith. 1976. *Access to Canadian government publications in canadian academic and public libraries.* Ottawa: Canadian Library Association.

Line, Maurice B. 1971. The information uses and needs of social scientists: an overview of INFROSS. *ASLIB Proceedings* 23: 412–34.

Matsuda, Yoshiro, and Setsuo Suoh. 1986. Demand for statistics in Japan: summary report. *International Forum on Information and Documentation* 11 (1): 29–33.

Meyriat, Jean. 1984. Use of information in science and research: social sciences. *International Forum on Information and Documentation* 9 (3): 10–12.

Morton, Bruce. 1995. Canadian federal government policy and Canada's electronic information industry. *Government Information Quarterly* 12 (3): 251–95.

Morton, Bruce, and Steven D. Zink. 1991. Contemporary Canadian federal information policy. *Canadian Public Administration* 34(2): 312–38.

Newson, Douglas. 1990. The role of a government agency in the dissemination of information: a perspective from Statistics Canada. *Canadian Journal of Information Science* 15(2): 74–79.

Nilsen, Kirsti. 1996. Survey of social scientists use of statistics. Unpublished. Faculty of Information Studies. University of Toronto.

Nilsen, Kirsti. 1993. Canadian government electronic information policy. *Government Information Quarterly* 10(2): 203–20.

Regan, Priscilla Marie. 1981. Public use of private information: a comparison of personal information policies in the United States and Britain. Ph.D. Dissertation. Cornell University.

Report of the Working Group on the Quality of Economic Statistics to the [President's] Economic Policy Council. 1987. April, 21. Cited in Wallman 1988, 11.

Ruus, Laine. 1987. Computer readable products from Statistics Canada: available but not affordable. Unpublished report. Vancouver: Data Library, University of British Columbia.

Slater, Margaret. 1989. *Information needs of social scientists.* British Library Research Paper 60. Boston Spa: British Library Research and Development Department.

Starr, Paul, and Robs Corson. 1987. who will have the numbers? The rise of the statistical services industry and the politics of public data. In *The Politics of numbers,* ed. William Alonso and Paul Starr, 415–47. New York: Russell Sage Foundation.

Statistics Canada. *Annual reports.* Ottawa: Statistics Canada.

Statistics Canada. *Estimates.* Ottawa: Supply and Services.

Wallman, Katherine K. 1988. *Losing count: the federal statistical system.* Washington: Population Reference Bureau.