Supermarket: Where Do Social Scientists Shop?

*by Kirsti Nilsen**

This paper presents some findings of research which examined the statistics and data sources used by Canadian social scientists, the formats in which they obtained the data, and their preferences with respect to data formats. Five disciplines were the focus of the research: economics, education, geography, political science, and sociology, based on a literature review which is summarized below. The research was part of a larger study which examined the effects of government information policy on Canadian social scientists. That research focused on policy- initiated price and format changes at Statistics Canada. (Nilsen, 1996, 1997, 1998). In order to monitor the effects of the policy it was necessary to determine which statistics and data sources were used and any changes in that use over a period before and after policy implementation. Using both bibliometric and survey methods to gather data, the study identified statistics sources used in published articles over the period 1982 to 1993, and supplemented those findings with a survey of authors in the Fall of 1995.

The terms "statistics" and "data" have unique definitions; however, for the purposes of this research, the terms tended to be used interchangeably as they are in everyday speech. In the survey, respondents were asked about their use of "statistical data (i.e. numeric information)".

Literature Review

Research on social scientists' use of and demand for materials has confirmed that social scientists do use statistics and raw data. Because governments collect, analyze and publish the largest amounts of data, social scientists will use government-produced statistics, along with other statistics sources. Obviously not all social science disciplines use published statistics and data sets to the same extent. In order to determine which disciplines should be the focus of this study, published research on social scientists' use and demand for materials was examined. It provided the data needed to identify those social science disciplines which use published statistics. Where statistics were not specifically identified, use of government publications served as an indicator of use of statistics because, as Hernon had shown, social scientists use government publications to obtain statistics more than for any other purpose (Hernon, 1979, p. 10). No research was found which distinguished between use of statistical publications of governments versus those of other publishers.

Use of Statistics by Social Scientists

The first major study of users of social science materials was undertaken by the Investigation into Information Requirements of the Social Sciences (INFROSS, 1971) in the United Kingdom. With 1,089 social science researchers responding to the INFROSS survey, it has been described as the largest, most ambitious and influential study in the area (Slater 1989, pp. 1). No research on a comparable scale has been done in North America.

INFROSS provided extensive data on the use of a variety of types and physical forms of information, along with data on information demand, by discipline, and with comparisons among disciplines. It specifically addressed the question of the use and perceived importance of statistics by researchers in each of the disciplines covered (anthropology, economics, education, geography, political science, psychology, and sociology).

The INFROSS study found that statistical, methodological and conceptual information was used by almost everyone, while historical and descriptive information was least used (Line, 1971, p. 416). Statistical material was used by 91% of respondents and over half used it frequently in their research. When asked to rate the importance of types of materials to themselves, INFROSS found that 58% of respondents rated statistical material as very important, 20% rated it as moderately important, and 12% rated it as not very important (INFROSS, 1971. vol.1, pp. 48, 50, 52).

With respect to disciplinary differences in use of statistical materials, INFROSS found that economists were the heaviest users of statistical data, followed closely by geographers. When asked to rate the importance of statistics, economists and geographers were much more likely than any other researchers to rate statistical material as "very important" and historians and anthropologists less likely to do so (INFROSS, 1971, vol.1, pp. 43, 51).

In its analyses of statistics use, INFROSS did not discriminate between data which were self-collected and

data gathered and published by someone other than the researchers themselves. However, the type of raw data used (e.g. interviews, experiments) was correlated with discipline of respondents (INFROSS, vol. 2, table 20). The report noted that psychologists were more likely to use empirically derived data from experiments conducted by themselves than were other social scientists (INFROSS, 1971, vol.1, p. 57).

Use of Government Information by Social Scientists

In reviewing the literature on citation studies, Hernon and Shepherd determined that the percentage of citations to government publications ranged from 2% to 36% (Hernon & Shepherd 1983, p. 227). Weech found that in various citation studies a median of 17.5% of total references were to government publications (Weech, 1978, p. 179).

The largest citation study of social scientists' use of materials was Design of Information Systems in the Social Sciences (DISISS, 1979), a follow-up study to INFROSS. DISISS collected data from 140 social science serials, published mostly in 1970, for an examination of social science literature via citation analysis. Out of 47,342 citations only 2.7% were to official (government) publications (DISISS, 1979, p. 75). The variability in the findings on use of government publications among social scientists can be accounted for by disciplinary differences in the choice of disciplines included in citation studies. Low percentages in general relate to the fact that statistical sources are often not cited in footnotes or reference lists (Hernon & Shepherd, 1983).

The INFROSS survey found that 34% of social science researchers used government publications "often", while 23% never used this form of material (INFROSS, 1971, vol.1, p. 53; Line, 1971, p. 417). When use of government publications was examined by discipline, the investigation found that 53% of researchers in economics stated that they sometimes or often used them, followed by those in sociology (41%), education (29%), geography (22%), and political science (20%). Fewer than 10% of researchers in anthropology, history and psychology used government publications (INFROSS, 1971, vol.2, table 59).

Hernon (1979) investigated the use of government publications by faculty members from economics, history, political science, and sociology departments in American colleges and universities. He found a statistically significant difference among the four disciplines in frequency of document use, with economists and political scientists as the heaviest users of government publications, which was consistent with the INFROSS findings (Hernon, 1979, pp. 9,45).

Some research has shown which disciplines seek statistical information within government publications. Hernon found that the "top priority of economists and sociologists [in using government publications] is to gather census and normative data," and that historians used government publications for historical data more, while political sciences use them equally for statistics and current events information (Hernon, 1979, p. 51). Other studies by Hernon and Shepherd (1983) and Hernon and Purcell (1982) corroborated Hernon's earlier findings.

Determining the Disciplines for This Research

On the basis of the INFROSS and DISISS research, which has been substantiated by other research, a typology of use of statistics and government publications was developed, as shown in Table 1.

TABLE 1	
	Typology of Use of Statistics and Government Publications: *By Discipline*
(1)	Disciplines which seldom or never use statistics: *Anthropology, History*
(2)	Disciplines which seldom or never use government publications: *Anthropology, History, Psychology*
(3)	Disciplines which sometimes or often use statistics: *Economics, Education, Geography, Political Science, Psychology, Sociology*[1]
(4)	Disciplines which use self-collected statistics: *Psychology*[2]
(5)	Disciplines which use primarily published statistics: *Economics, Education, Geography, Political Science, Sociology*
(6)	Disciplines which sometimes or often use government publications: *Economics, Education, Geography, Political Science, Sociology*[3]

Based on this typology, and the research which supports it, five disciplines were identified which use primarily published statistics *and* sometimes or often use government publications. These five disciplines were economics, education, geography, political science, and sociology. Thus, these five disciplines defined the domain of this research.

Methodology

Two methods were used to gather data on the use of statistics sources. Bibliometric analysis provided objective evidence of use of statistics, while a survey supplemented the findings with more subjective data. A systematic,

stratified and proportionate sample of 360 articles was selected from a population of 5,414 articles in 21 Canadian social science journals in the five disciplines noted above. The source journals were published in English or French in Canada, covered primarily Canadian topics, focussed widely in the discipline, were peer-reviewed, and published over the entire period 1982 to 1993. All journals which met these criteria were included. Articles to be included in the population to be sampled were those listed in the tables of contents under "Articles" or "Research Notes" or similar headings. In the final sample, the disciplines were represented in proportion to the amount of publishing in the 21 journals: economics 26.9% (97 articles), education 22.5% (81 articles), geography 7.2% (26 articles), political science 18.1% (65 articles), and sociology 25.3% (91 articles).

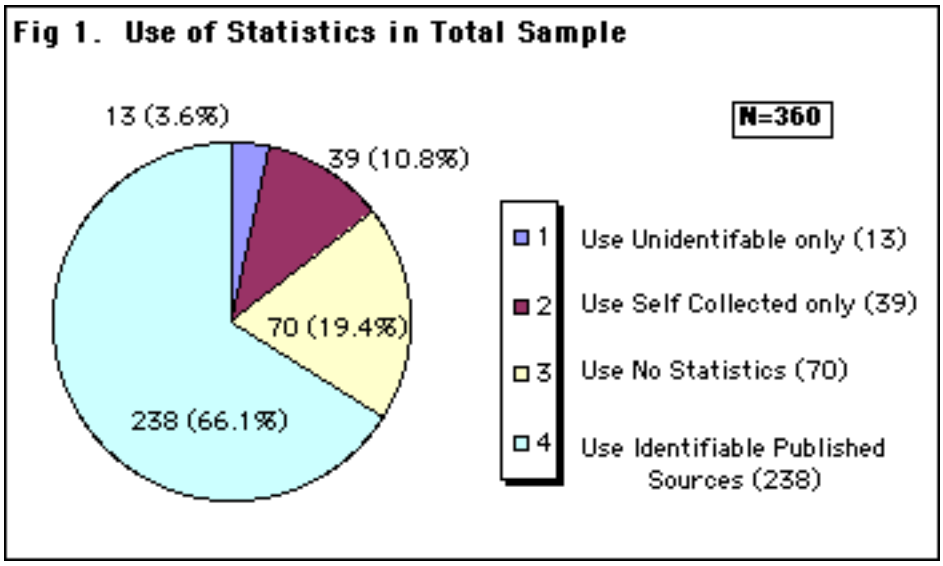
The 360 articles in the sample were examined and data were collected from the text, tables, and citations. The bibliometric examination revealed the statistics sources used by the authors of the articles. All uses of statistics sources, whether documented or not, were recorded, whether governmental, nongovernmental, Canadian or foreign. More detailed information about the use of Statistics Canada was gathered for the policy effects aspect of the larger study. Data analysis dealt with the complete sample and, in more detail, with a subset of 207 articles which were identified as using published statistics *and* written with a Canadian focus or setting.

A survey questionnaire was sent in English or French to 163 authors (all who could be located) of these 207 articles. Ninety-seven responded (59.5%). The questionnaire asked for background information, extent of use of statistics, statistics sources used, formats used and preferred, means of obtaining data, and opinions regarding prices and formats of data..

Findings

The 360 articles sampled for the bibliometric component of the research were categorized as to discipline, type, geographical focus or setting (if any), and language. The categorization was by discipline of the journal, (which was not necessarily the discipline of the author or of the subject covered). Most articles (78.4%) could be categorized by type as either empirical (200, 55.6%) or descriptive (82, 22.8%), both of which were likely to use statistics. The remaining 78 articles (21.6%) were either historical, opinion, methodological, or theoretical, articles less likely to use statistics. The geographical focus or setting was Canadian in 269 of the articles (74.7%), and the focus was not Canadian in 34 articles (9.5%). The remaining 57 articles (15.8%) could not be categorized geographically, usually because of their methodological or theoretical focus. Two-thirds of the articles (239, 66.4%) were in English, 121 articles (33.5%) were in French.

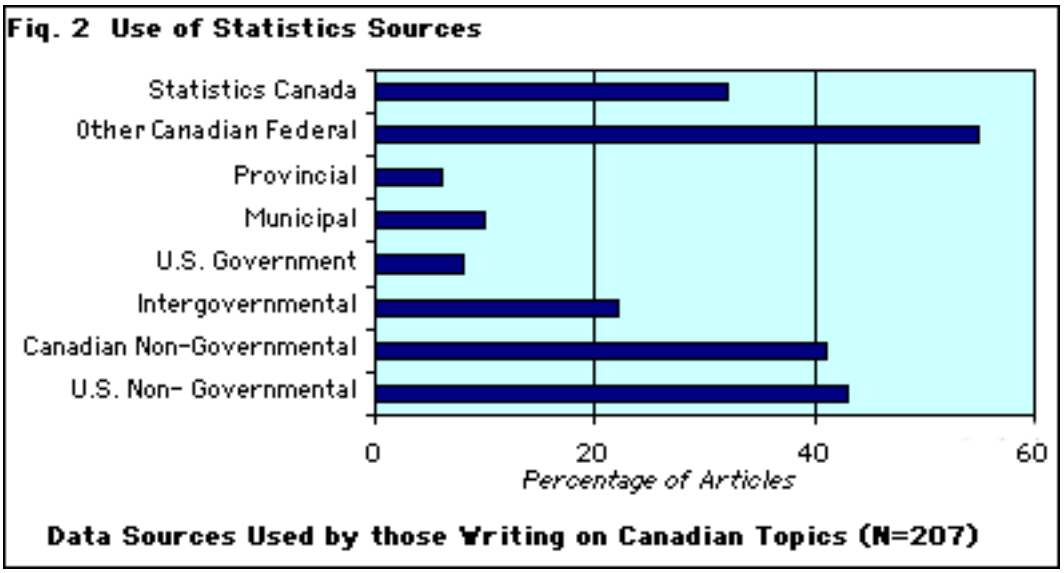
As expected, not all of the articles used statistics. As Figure 1 illustrates, 70 articles (19.4%) made no use of any statistics, most of these were categorized as theoretical or methodological. Thirty-nine articles (10.8%) used only self-collected data derived by the author from experiments or other research methods. A few articles (13, 3.5% used only unidentifiable published statistics which could not be categorized as to source. The remaining 238 articles (66.1%) used identifiable published statistics. More than 70% of the articles in each discipline (excepting education) used identifiable published statistics. Some of these also used self-collected data.



A subset of 207 articles was identified which had a Canadian focus or setting and used published statistics and this subset provided the data which follows.

Statistics Sources Used

Information was gathered on the use of the following broad categories of statistics sources: Statistics Canada, other Canadian federal and provincial/municipal governments, foreign governments, intergovernmental, nongovernmental. More detailed information was gathered on the use of Statistics Canada in terms of formats used. It was found that social scientists used a wide variety of statistics sources and many used multiple sources. **Figure 2** illustrates the percentage of articles which used the various



sources.

As can be seen, Statistics Canada was used by 41.1% of the articles, and other Canadian federal sources were used by an almost equal number (40.6%**).** American sources (governmental and nongovernmental combined) were used almost as much as was Statistics Canada, which was somewhat surprising in these 207 articles with a Canadian focus or setting. Nongovernmental sources were used by the highest percentage of articles (71%). Nongovernmental sources include trade and scholarly books and journals, associations, universities, business, think tanks, polling organizations, etc.

The survey respondents indicated higher use of all statistics sources (except nongovernmental) than was found in the bibliometric research. This probably results from the fact that the bibliometric analysis was looking at one-time use in a single article while the survey questioned life-time use. For example, 86.5% indicated that they had used Statistics Canada at some time, but only 41.1% used Statistics Canada in the articles. However, only 41.5% of survey respondents indicated that they used Statistics Canada often or almost always (i.e. more than 50% of the time in the years between 1985 and 1995), which is more consistent with the bibliometric finding. They also indicated less use of nongovernmental sources than was found in the bibliometric analysis. This difference might result because respondents might have been thinking of major nongovernmental suppliers such as polling organizations, rather than their use of sources from which they might obtain single facts such as a book or journal article.

Disciplinary differences

There were statistically significant disciplinary differences in use of most statistics sources in the articles as is shown in Table 2. Note that variation by discipline was statistically significant for all sources but provincial government and foreign government sources. Table 2 shows the importance of other Canadian federal government sources in articles from economics and political science journals. Those writing in education and political science journals were more likely to use other Canadian federal government sources more than they used Statistics Canada. Geography and sociology journal articles used Statistics Canada more often, and economics used the agency's

Table 2
Statistics Sources Used: By Percent of Discipline in Articles With a Canadian Focus or Setting n=207)

Statistics Source	Eco	Edu	Geo	Pol	Soc
Statistics Canada (p = .001)	58.3	23.3	72.2	19.2	42.1
Other Cdn. Federal (p = .001)	58.2	30.0	22.2	51.1	26.3
Provincial (p = .128)	12.7	30.0	22.2	27.7	33.3
Municipal/Regional (p = .025)	5.5	13.3	27.8	4.3	5.0
Foreign Govt (p = .086)	21.8	6.7	5.6	10.6	7.1
US Federal Govt (p = .243)	18.2	6.7	5.6	8.5	7.1
Intergovernmental (p = .165)	7.3	3.3	5.6	10.6	0.0
Nongovernmental (p = .018)	67.3	73.3	56.6	89.3	63.2
Canadian Nongovt (p = .003)	50.9	53.3	55.6	80.9	43.9
American Nongovt (p = .028)	29.1	20.0	22.2	51.1	29.8

statistics at approximately the same rate as they use the statistics of other federal sources. Nongovernmental sources were important for all disciplines, particularly political science, which made heaviest use of both Canadian and American nongovernmental sources.

Use of Computer Readable Products

In the survey (conducted in the Fall of 1995), most respondents (81%) indicated that they had used computer readable formats at some time. There was statistically significant variation by discipline in these responses, with 100% of those who had published in economics and geography journals indicating prior use, while 77.8% of those in sociology, 73.7% of those in political science, and 56.3% of those in education indicating such use. However, when asked how they normally obtained statistics most still used paper (print) formats more than computer readable files. Of the 97 respondents, 74 (76.3%) indicated that they obtained data in paper format, while 59 (60.8%) used computer readable formats, or both formats, as seen in Table 3.

TABLE 3
Formats in Which Survey Respondents Normally Obtained Statistics (N=97)

Formats	No	Percent*
Print (i.e., paper)	74	76.3
Microform (microfiche, microfilm)	13	13.4
Computer readable files**	59	60.8
Special tabulations***	36	37.1
Collect my own data	60	61.9

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

Questioned as to how they normally acquired the data they used, responses are shown in Table 4:

Respondents were then asked to rank their first preference and their first three preferences of the various means of acquiring data, as shown in Table 5.

Note that where 47% in Table 4 used a library to acquire

TABLE 4
How Survey Respondents Normally Acquired Statistics (N=97)

Means of Acquiring Statistics	*No*	*Percent**
Collect My Own Data	56	57.7
Purchase Computer Readable Files	50	51.5
Use a Library for Paper or Microform	46	47.4
Use a Data Library	37	38.1
Purchase Paper Copies	35	36.1
Purchase Special Tabulations	30	30.9
Use the Internet	19	19.6
Use a Departmental Collection for Computer Readable Data	15	15.5
Use a Departmental Collection for Paper Copies	10	10.3
Purchase Microform Copies	9	9.3
Other	2	2.1

* Totals exceed 100% because respondents could indicate all means of data acquisition which they used

paper copies, for only 20% was that a top three preference. A larger percentage ranked purchasing computer readable files as a top three preference than had indicated normally acquiring data in this way. Also, more preferred to use a data library. Fewer preferred to collect their own data than actually did so.

The bibliometric analysis provided objective data on the actual use of paper and computer readable formats. The determination of use of products by format focussed on the 85 articles with a Canadian focus or setting which used Statistics Canada as a statistics source. Using various Statistics Canada catalogues and other sources where necessary, the researcher determined the formats of the Statistics Canada issues used in the articles, if the author had not provided this information. Seventy-one (84%) of these 85 articles used paper issues, while 29 (34%) used computer readable “issues”, with some articles using both formats. The format of some issues could not be determined in 11 articles. The ratio of number of articles using paper issues to the number using computer readable issues was 2.5:1. Table 6 illustrates variation in the number of articles using issues by format.

TABLE 5
Ranked Preferences for Acquiring Statistics (N=79)

Means of Acquiring Statistics	Percent Indicating First Choice	Percent Indicating One of Top Three Choices
Purchase Computer Readable Files	22.8	68.4
Collect Own Data	22.8	68.4
Use a Data Library	20.3	51.9
Purchase Paper Copies	10.1	26.6
Use a Library for Paper Copies	10.1	20.3
Purchase Special Tabulations	5.1	19.0
Use the Internet	2.5	22.8
Use a Departmental Collection for Computer Readable Files	2.5	16.5
Use a Departmental Collection for Paper Copies	1.3	6.3

Variation over the two time periods 1982-1987 and 1988-1993 in the number of articles using these formats was statistically significant for paper issues.

It should be noted that while the year-to-year variations are not statistically significant, the percentage of articles in the sample, which used computer readable formats as a statistics (or raw data) source, increased in the last two years studied. These formats were used in 20% of the articles in 1992 and 36% of the articles in 1993. **Figure 3** illustrates year to year variation.

This might be an indication of a trend which might have been evident in a larger sample and which could be examined in further research.

TABLE 6
Number of Articles Using Statistics Canada Products: By Format
In Articles Which Used Statistics Canada (N=85)

Time Period	Paper (p =.045)	Computer Readable (p =.290)
1982-1987 (n=39)	36 (92%)	11 (28%)
1988-1993 (n=46)	35 (76%)	11 (28%)

There were statistically significant disciplinary differences in use of Statistics Canada's paper ((r = .009) and computer readable formats (r =.014) among the 207 articles written with a Canadian focus or setting. These are illustrated in **Figure 4**.

These data apply only to Statistics Canada formats, and as the discussion below indicates, other sources of computer readable information were used by these disciplines as well. Political science for example, showed little use of Statistics Canada overall, but was a heavy user of nongovernmental materials, and made some use of computer readable sources, such as the National Election Studies.

The decline in the number of articles which used paper formats might be attributed to the declining publication of

Figure 3
Use of Statistics Canada Paper and Computer Readable Products
In Articles Written on Canadian Topics (N=207)

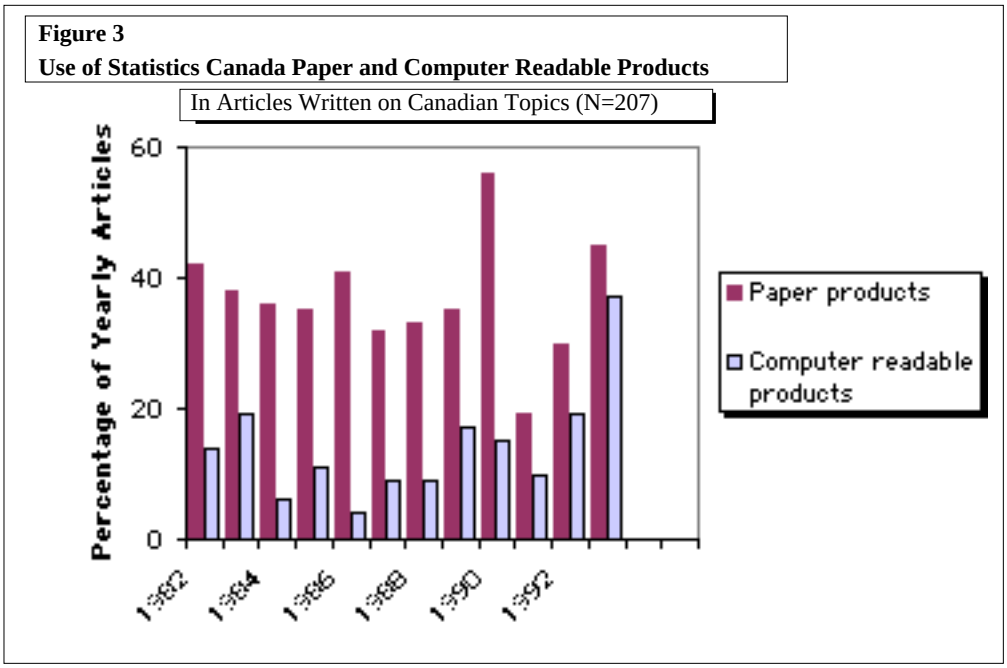


Figure 4: Use of Paper & Computer Readable Formats
Percent of articles in each discipline using Statistics Canada materials in each format

paper formats at Statistics Canada, rather than any absolute preference. However, the survey responses suggest that computer readable formats are preferred. It should also be noted that of the surveyed respondents who began to do research after 1980 (younger researchers?) 85% indicated that they had use Statistics Canada computer readable files at some time, while of those who began to do research before 1970, only 46.5% had used them. This suggests that in the future data users will rely on the computer readable files to an ever greater extent.

Machine Readable Data Files Used

When computer readable files were used as major sources of data in articles, the titles of the MRDFs were recorded. Because authors tended to cite these materials incompletely, if at all, the following discussion should be interpreted cautiously.

Statistics Canada computer readable files were used by articles in education, economics, geography and sociology, with articles in economics journals using the greatest variety of files. Special tabulations were used by economics and geography authors for census data, and by economics authors for family expenditures, manpower, manufacturing, and agriculture data. An education article used the Labour Market Activity file; a Justice database was used by one sociology article. Public Use Sample Tapes were used by one article from economics and two from sociology. CANSIM was mentioned by only one author.

Other Canadian federal MRDFs were used in economics and political science articles. Three economics articles used Labour Canada files, and the International Trade Data Bank was used by a political science article. Quebec provincial health databases were used by two sociology articles. One US government database was used by an economics article (Dept. of Agriculture CRIS), and two sociology articles used US government data obtained from ICPSR. Canadian universities were an important source for data for sociology articles, and to a lesser extent for political science and economics articles. Here, the National Election Studies were used by an article in economics and one in political science. Both York University's Quality of Life Survey and the University of Western Ontario's Canadian Fertility Study were each cited by one economics article and one sociology article. Two francophone sociology journals cited SOREP data on Quebec population, while a third cited a database created at the École des hauts études commerciales. Other databases used include one use of the FAO trade tape, one use of the data from the Correlates of War Project (US university), and proprietary databases were cited by one economics article.

The above information suggests a rather limited used of MRDFs by Canadian social scientists. However, as noted above, authors do not cite these sources with any consistency. Additionally unless an item could be clearly identified as an electronic file, it was assumed to be a paper product if such a product was available in print. Thus, it is possible that some items that were recorded as paper products were in fact electronic files. Bibliometric analysis of the use of electronic files suffers from inconsistencies in citation practices. This was noted as early as 1982 (White), but the situation has not improved.

Discussion

The findings of this research are consistent with the findings of earlier studies cited in the literature review. Social scientists do indeed use a wide variety of sources to obtain statistics and raw data. There are statistically significant variations in the sources used among disciplines. If any agency such as Statistics Canada wishes to expand its market, analyses by discipline can assist in identifying target consumers, or areas where its products are not meeting the needs of researchers.

At the time period covered by the bibliometric research (1982-1993) and the survey (1995), social scientists still used paper products more than computer readable products to obtain statistics and data, but there was a statistically significant decline in the use of paper products.

Additionally, respondents to the survey were enthusiastic about computer readable formats. These finding suggested that computer readable formats would be used more heavily in the future, Indeed, in 1999, we see much more availability of information on computers. It is highly likely that future research will show a much stronger shift to electronic formats for data access. The findings of this research can provide baseline data for future comparisons.

Bibliography

Brittain, J.M. 1970. *Information and Its Users: A Review with Special Reference to the Social Sciences*. New York: Wiley, 1970.

[DISISS]. 1979. Design of Information Systems in the Social Sciences. *The Structure of Social Science Literature as Shown by Citations*. Research Report A no.3. Bath UK: Bath University Library.

Hernon, Peter. 1979. *Use of Government Publications by Social Scientists*. Norwood NJ: Ablex.

Hernon, Peter, and Clayton A. Shepherd. 1983. "Government Publications Represented in the *Social Sciences Citation Index*: An Exploratory Study." *Government Publications Review* 10: 227-244.

Hernon, Peter, and Gary R. Purcell. 1982. "Document Use Patterns of Academic Economists." Chapter 5 in their *Developing Collections of US Government Publications.* Greenwich CT: JAI Press.

[INFROSS]. 1971. Investigation into Information Requirements of the Social Sciences. *Information Requirements of Researchers in the Social Sciences*. Research Report no.1. 2 vols. [Bath UK]: Bath University Library.

Line, Maurice B. 1971. "The Information Uses and Needs of Social Scientists: An Overview of INFROSS." *ASLIB Proceedings* 23: 412-434.

Nilsen, Kirsti. 1996. "The Effects of Electronic Publication on Social Science Researchers in Canada." In: *Electronic Publishing: Its Impact on Publishing, Education and Reading*. pp. 1-26. Proceedings of the 24th Annual Conference, Canadian Association for Information Science. Toronto, 2-3 June, 1996. Toronto: CAIS.

Nilsen, Kirsti. 1997. Social Science Research in Canada and Federal Government Information Policy: The Case of Statistics Canada. Unpublished doctoral dissertation. University of Toronto. UMI DANQ28027

Nilsen, Kirsti. 1998. "Social Science Research in Canada and Government Information Policy: the Statistics Canada Example." *Library and information Science Research* 20 (3): 211-234.

Slater, Margaret. 1989. *Information Needs of Social Scientists*. Boston Spa UK: British Library Research and Development Department.

Weech, Terry L. 1978. "The Use of Government Publications: A Selected Review of the Literature." *Government Publications Review* 5: 177-184.

White, Howard. 1982. "Citation Analysis of Data File Use." *Library Trends* 30: 467-477.

Footnotes:

[1] Based on INFROSS results and including only those in which 60% or more of respondents indicated that they use statistics.

[2] Economists, social geographers, sociologists and some political scientists also use experimentally derived data, but to a lesser extent.

[3] Based on INFROSS results and including those in which 20% or more of researchers claim they use government publications.

* Paper presented at the IASSIST Conference, May 19, 1999, Ryerson Polytechnic University, Toronto, Ontario..

Kirsti Nilsen, Faculty of information and Media Studies,University of Western Ontario, London, Ontario N6A 5B7, Canada