
Integrating Data Resources and Library Resources: The SPIRES Experience

by Slavko Manojlovich
Memorial University of Newfoundland
St. John's, Newfoundland

INTRODUCTION

The CANSIM database, maintained by Statistics Canada, contains over 400,000 socio-economic time series. The CANSIM University Base, a subset of the main CANSIM base, contains approximately 35,000 numeric time series and is made available to academic institutions on magnetic tape for instructional and research purposes. Prior to 1991, access to the CANSIM University Base at Memorial was provided by the Economics Department via a locally developed FORTRAN program which retrieved time series from tape for specified databank numbers. In 1991 the Library, for budgetary and other reasons, took over the CANSIM subscription. CANSIM has since been loaded into SPIRES, the library's database management system, which is also used to maintain the library catalogue and other library and departmental databases. CANSIM is now accessible to the university community using the same interface as other library databases. It has also been linked to the library catalogue for the display of holdings information associated with related print publications. The remainder of the article will describe various issues regarding loading a numeric database in a library environment.

SELECTING A DATABASE MANAGEMENT SYSTEM

The SPIRES database management system, developed by Stanford University, has been used for over 15 years to manage a variety of information resources including bibliographic, numeric, full-text and image databases. Memorial has been using SPIRES for several years to maintain its library catalogue. Various components of an integrated library system including acquisitions and cataloguing were developed locally, whereas, circulation was obtained from Rensselaer Polytechnic Institute, a member of the SPIRES Consortium. In addition to the library catalogue Memorial has mounted a number of locally developed and commercial databases using the FOLIO interface available in SPIRES (Screen No. 1).

SPIRES has a powerful set of development tools which enable you to tailor the system to accommodate a variety of data types. Since CANSIM was the first numeric database which the library was loading into SPIRES an important first step in the loading process was to identify the DBMS functionality which was required in order to support adequate access to a numeric database. The

following functions, although not critical for bibliographic database support, were important for accessing CANSIM:

- 1) Since end-users typically retrieve time series data for specified time periods SPIRES must prompt for the time series start and end dates.
- 2) SPIRES must format the data for tabular display or for input to statistical analysis software.
- 3) SPIRES must output data to a file on the mainframe or on the end-user's microcomputer.

THE CANSIM DATABASE IN SPIRES

The CANSIM file supplied to Memorial contains 1.5 million card images. The logical record describing one time series is comprised of various fixed fields (codes and text strings) and a variable number of data values (Figure 1). SPIRES provides a utility for loading data in its original form thereby alleviating the need to write loader programs. No problems were encountered in loading the CANSIM file. SPIRES has a library of processing functions which allows you to read in data in any form, store it as you like and output it in any form. This inherent flexibility of SPIRES made it easy to implement prompting for start and end dates and the various output formats. SPIRES was also able to accommodate the requirement to output a variable number of data values for specific years when retrieving weekly time series. The various features of CANSIM in SPIRES are illustrated in the following sample search session (please refer to screen displays at the end of the article):

- 1) A variety of **HELP** screens navigate users through a search. (Screen No. 2).
- 2) Users can search CANSIM directly using the **FIND** command on various indexes (Screen No. 3). Terms can be combined from various indexes using boolean operators. The **BRIEF** record display output from a keyword search on "NEWFOUNDLAND AND WOMEN AND UNEMPLOY#" is illustrated in Screen No. 4.
- 3) The **FULL** record display (Screen No. 5) includes in

the **SOURCE** field not only source publication information as supplied by CANSIM but also information regarding Memorial's holdings for that title. For each displayed record SPIRES looks up the holdings information in the library catalogue and includes it in the source field. The addition of holdings information enables the user to easily consult the corresponding print publication for additional information describing the time series or to obtain older data not included in CANSIM. SPIRES performs a corresponding look-up for users searching the library catalogue and notifies them of related information in the CANSIM file (see the **NOTES** field in the record display from the library catalogue — Screen No. 6).

4) Users have the option of displaying data in tabular form (Screen No. 7) or as raw data formatted for input to a variety of time series analysis software packages (Screen No. 8). The user selects the desired format by issuing a **DIS TABLE** or **DIS DATA** command. SPIRES will then prompt the user for the time series start and end dates. Output in either tabular or raw data format may be directed to a file on the mainframe or to the user's microcomputer by issuing either the **SAVE TABLE** or **SAVE DATA** commands. In **SAVE** mode SPIRES prompts the user for the output file name.

5) Users who are uncertain of the appropriate search terms may use the **SPIRES BROWSE** command to scan entries in various indexes (Screen No. 9 and No. 10). Users can selectively display time series from the hit list of displayed terms. All text indexes are may be browsed in the CANSIM application.

6) The above session used SPIRES menu-driven Folio **GUIDED** mode. A **COMMAND** mode option is also available for more experienced users.

THE INTEGRATING ROLE OF SPIRES

The concept of integration in library automation literature typically refers to linkages between various modules in an integrated library system. The CANSIM application described above illustrates how SPIRES expands upon the traditional meaning of integration. Integration in a SPIRES environment also includes:

1) **Interface Integration:** the use of a common user interface for accessing a variety of databases.

2) **Database Integration:** the provision of automatic linkages between databases thereby expanding the user's knowledge base (eg. source publication link between CANSIM and the library catalogue).

3) **System Integration:** the output of data in a variety of

formats for direct input to other systems (eg. SAS).

4) **Workstation Integration:** support for saving data to a file on the mainframe or on the user's workstation.

All of the above further integrate the user into his/her information environment.

CD-ROM VERSION OF CANSIM

Statistics Canada is also distributing a larger portion of the CANSIM database in CD-ROM format. Although CD-ROM is an excellent cost-effective medium for the distribution of large quantities of data it falls short in terms of "integration" as described above, especially in a university environment. The CD-ROM medium forces the user to learn a new user interface. It is not directly linkable to the library catalogue and other campus databases. Access to the database may be restricted to a single workstation. If the CD-ROM is mounted on a campus network, access may be restricted to users with particular hardware. Remote access to a CD-ROM by an end-user working on an old VT-100 terminal may not be possible at all. Migrating data from the CD-ROM to the end-user's statistical analysis software (which is probably on a mini or mainframe) may also be quite cumbersome.

CONCLUSION

Thanks to the integrating power of SPIRES, access to CANSIM at Memorial University is the same as accessing the library catalogue or any other bibliographic database. Based on the success of mounting CANSIM, Memorial is planning on loading census and other fact databases in SPIRES thereby further expanding access to the world of numeric data resources for the library user.

NOTES

1. Additional information on SPIRES is available from: SPIRES Consortium Office, Stanford University, Stanford, California.

¹ Presented at the IASSIST 91 Conference held in Edmonton, Alberta, Canada. May 14 - 17, 1991. Slavko Manojlovich, Assistant to the University Librarian for Systems and Planning, Memorial University of Newfoundland, St. John's, Newfoundland.

Figure 1

SAMPLE CANSIM RECORD AS SUPPLIED BY STATISTICS CANADA

ADD B 1 90-11-02 195319901210 5 2 310 0 1 199999999*
 (12X, 4F17.0)
 BANK OF CANADA ASSETS AND LIABILITIES, WEEKLY SERIES (101-103) AND MONTHLY SERIES
 (1-3), WEDNESDAYS AND AVERAGE OF WEDNESDAYS, UNADJUSTED, MILLIONS OF DOLLARS.
 B.OF C-STATEMENT/AVE TOTAL ASSETS DOLLARS SCALAR FACTOR 06
 SOURCE BANK OF CANADA REVIEW
 AVERAGE OF WEDNESDAYS CANSIM SERIES IDENTIFIER 000911.1
 NOTE DATA PUBLISHED IN THE BANK OF CANADA REVIEW APPROXIMATELY 30
 CALENDAR DAYS AFTER END OF REFERENCE MONTH.

B	1	1	2348.	2318.	2332.	2352.
B	1	2	2354.	2352.	2410.	2408.
B	1	3	2371.	2364.	2429.	2444.
B	1	4	2390.	2404.	2355.	2357.
B	1	5	2427.	2431.	2309.	2284.
B	1	6	2326.	2304.	2382.	2420.
B	1	7	2369.	2226.	2278.	2310.
B	1	8	2316.	2357.	2433.	2468.
B	1	9	2464.	2476.	2532.	2547.
B	1	10	2509.	2368.	2421.	2472.
B	1	11	2467.	2511.	2528.	2531.
B	1	12	2519.	2543.	2550.	2571.
B	1	13	2514.	2406.	2429.	2492.
B	1	14	2519.	2580.	2604.	2629.
B	1	15	2632.	2645.	2696.	2670.
B	1	16	2606.	2540.	2574.	2646.
B	1	17	2652.	2719.	2800.	2855.
B	1	18	2885.	2997.	2956.	2951.
B	1	19	2800.	2753.	2768.	2809.
B	1	20	2838.	2857.	2857.	2928.
B	1	21	2880.	2848.	2943.	2869.
B	1	22	2822.	2728.	2736.	2816.
B	1	23	2830.	2842.	2902.	2905.
B	1	24	2860.	2895.	2950.	2927.
B	1	25	2906.	2824.	2876.	2896.
B	1	26	2920.	2909.	2981.	2998.
B	1	27	3030.	3066.	3064.	3066.
B	1	28	3062.	2940.	2990.	3075.
B	1	29	3105.	3227.	3242.	3309.
B	1	30	3178.	3205.	3215.	3221.
B	1	31	3136.	3012.	3072.	3167.

.....

Screen No. 1
MENU OF PUBLIC ACCESS DATABASES AT MEMORIAL

Folio contains 17 files. Press the RETURN key to see the rest of the list.

Public Information Files

1. BIBLIO	MUN Library Online Catalogue
2. CANSIM	Canadian Socio-Economic Time Series Database
3. CNS ARCHIVES	Archival records of the Centre for Nfld. Studies
4. EXTENSION	Division of Extension Resource Library
5. LABBIB	Canadian Labour Bibliography
6. LINSIC	Labrador Institute of Northern Studies Info Cen.
7. MICROLOG	Canadian Research and Report Literature
8. MUNFLA	MUN Folklore and Language Archive
9. OEIC	Ocean Engineering Information Centre
10. PAB	Newfoundland Periodical Article Bibliography
11. SLIDES	Grenfell College Fine Arts Slides Database

==> Press the RETURN key to see the rest of the list,
or select a file by typing its name or number

WHICH

FILE?

CANSIM

Screen No. 2
INITIAL CANSIM HELP SCREEN

-CANSIM selected

CANSIM is Statistics Canada's computerized data bank and information and retrieval service. The CANSIM University Base on FOLIO contains 32,425 of the most popular time series in the main base. Subject areas covered by CANSIM include system of national accounts, population, labour, prices and international trade. The database is currently updated once a year (the last update was December, 1990). Time series may be output as either tabular displays or in raw data format suitable for input to a variety of time series analysis software packages. Raw data or tables may be output to a file at NLCS or downloaded to the user's microcomputer using KERMIT communications software. The CANSIM FULL display includes the MUN library holdings of the associated print publications. Enter HELP CANSIM OUTPUT for more information on CANSIM display formats. For additional information contact Joy Tillotson, Information Services, QE II library, ext. 7427.

For more information on this file: type HELP.
To search this file: type FIND.
To search headings in sequence: type BROWSE.
To select a different file: type SELECT.
To see all your options: type OPTIONS.
YOUR RESPONSE: FIND

Screen No. 3
MENU OF CANSIM INDEXES

You can search the CANSIM file for any of the following information:

Type of Search	Example
MATRIX matrix numbe	911 or 000911
SOURCE keyword index to source publications	BANK OF CANADA REVIEW
MAWORD keyword index to matrix titles	BANK OF CANADA ASSETS
SEWORD keyword index to series titles	mortgage#
KEYWORD keyword index to titles and notes	consumer price
FREQ frequency	MONTHLY
DB databank numbers	B10

Indicate below the type of search you want by typing the name or names of the type of information you have, e.g. MATRIX.

Use the BROWSE command to examine entries in various indexes.

TYPE OF SEARCH: FIND WORD NEWFOUNDLAND WOMEN UNEMPLOY#

Screen No. 4
CANSIM BRIEF RECORD DISPLAY

CANSIM / Search: Find KEYWORD NEWFOUNDLAND WOMEN UNEMPLOY#

Result: 6 series

1)

Series: UNEMPLOYMENT RATE WOMEN 20-24 YRS. (monthly, 1975-1990) (Databank No: D774089) Matrix: NEWFOUNDLAND, BASIC LABOUR FORCE CHARACTERISTICS, MONTHLY FROM JAN 75, UNADJUSTED (RAW) IN THOUSANDS. SELECTED SERIES ARE LINKED TO PREVIOUS SURVEYS OF JAN 66 OR JAN 70. (No: 002078)

2)

Series: UNEMPLOYMENT RATE, 25 YRS AND OVER, WOMEN (monthly, 1975-1990) (Databank No: D772686) Matrix: NEWFOUNDLAND, BASIC LABOUR FORCE CHARACTERISTICS, MONTHLY FROM JAN 75, UNADJUSTED (RAW) IN THOUSANDS. SELECTED SERIES ARE LINKED TO PREVIOUS SURVEYS OF JAN 66 OR JAN 70. (No: 002078)

Series continue; press RETURN to see next page

To see a full series: type DISPLAY FULL followed by a number.

To begin a new search: type FIND or BROWSE.

To select a different file: type SELECT.

To get more information: type HELP or OPTIONS.

YOUR RESPONSE: DF 1

Screen No. 5
CANSIM FULL RECORD DISPLAY

CANSIM / Search: Find KEYWORD NEWFOUNDLAND WOMEN UNEMPLOY#
Result: 6 series
Series 1
Series: UNEMPLOYMENT RATE WOMEN 20-24 YRS. (monthly, 1975-1990) (Databank No: D774089)
Matrix: NEWFOUNDLAND, BASIC LABOUR FORCE CHARACTERISTICS, MONTHLY FROM JAN 75, UNADJUSTED (RAW) IN THOUSANDS. SELECTED SERIES ARE LINKED TO PREVIOUS SURVEYS OF JAN 66 OR JAN 70. (Matrix No: 002078)
Source: MONTHLY LABOUR FORCE DATA (71-001), STC
MUN HOLDINGS: The labour force. La population active 71-001, LOCATION: MUGD, HOLDINGS: v. [10-11]-[33]- 1954-
Notes: SOURCE MATERIALS COVERING BACKGROUNDS OF PREVIOUS LFS REVISIONS AND MODIFICATIONS OF DEFINITIONS, OF CONCEPTS AND OF LFS DESIGN, AS THEY ARE INTRODUCED IN THE CURRENT REVISION MAY BE OBTAINED FROM THE DIVISION. REQUESTS SHOULD REFER TO 1) 'LABOUR FORCE INFORMATION', CAT. NO. 71-001P, FEB76 2) 'RESEARCH PAPER' #2 AND #3 3) 'METHODOLOGY OF THE CANADIAN LABOUR FORCE', STATISTICS CANADA, CAT. NO. 71-526, OTTAWA, 1976
Scalar Factor: 00 Data Output Format: (10X, 4F17.1)
Missing Values = 9999999999. or equiv. Secure data = all asterisks.

Screen No. 6
LIBRARY CATALOGUE DATABASE RECORD DISPLAY

BIBLIO / Search: Find TITLE LABOUR FORCE
Result: 16 titles
Title 16
TITLE: The labour force. La population active
FORMER TITLE: Labour force bulletin. Main d'oeuvre bulletin 1945-19
PUBLISHED: Ottawa, Statistics Canada, Labour Force Survey Division
DESCRIPTION: v
DATES: v. [1]- 1945-
ISSN: 03806804
NOTES: V.32, no. 1-2 not published
Vol. numbering begins with v.6, no.1, Mar. 1960
Issues for 1945-49 called no.1-13
NOTES: Vols. for 1945-1971 issued by the Dominion Bureau of Statistics. ; 197 -19 by Statistics Canada, Labour Force Surveys Section
SUBJECT(S): Labor supply-Canada-Statistics-Periodicals.
OTHER ENTRIES: Statistics Canada. Labour Force Surveys Section
Canada. Dominion Bureau of Statistics
OTHER ENTRIES: Labour force bulletin
CALL NUMBER: 71-001, LOCATION: MUGD, HOLDINGS: v. [10-11]-[33]- 1954-
RSN: 75348973
==> NOTES: *** NOTE: Numeric data associated with this record are

Screen No. 7
CANSIM TABULAR RECORD DISPLAY

1)
Series: UNEMPLOYMENT RATE WOMEN 20-24 YRS. (monthly, 1975-1990) (Databank No: D774089)

Matrix: NEWFOUNDLAND, BASIC LABOUR FORCE CHARACTERISTICS, MONTHLY FROM JAN 75, UNADJUSTED (RAW) IN THOUSANDS. SELECTED SERIES ARE LINKED TO PREVIOUS SURVEYS OF JAN 66 OR JAN 70. (No: 002078)

Year	Jan	Feb	Mar	Apr
1985	32.7	30.0	29.2	31.0
Year	May	Jun	Jul	Aug
1985	32.7	33.8	25.8	27.0
Year	Sep	Oct	Nov	Dec
1985	25.4	26.0	25.9	23.3
			
			
			

Screen No. 8
CANSIM RAW DATA OUTPUT FORMAT

Series: UNEMPLOYMENT RATE WOMEN 20-24 YRS.
Matrix: NEWFOUNDLAND, BASIC LABOUR FORCE CHARACTERISTICS, MONTHLY FROM JAN 75, UNADJUSTED (RAW) IN THOUSANDS. SELECTED SERIES ARE LINKED TO PREVIOUS SURVEYS OF JAN 66 OR JAN 70. (Matrix No: 002078)

(10X, 4F17.1) (monthly, 1985-1990, Scalar Factor: 00)

	32.7	30.0	29.2	31.0
	32.7	33.8	25.8	27.0
	25.4	26.0	25.9	23.3
	26.9	25.1	28.0	26.1
	25.9	25.9	30.1	23.8
			
			
			

Screen No. 9
MENU OF CANSIM BROWSE INDEXES

You can browse the following types of information in the CANSIM file:

Type of Browse	Example
MATI matrix titles	Bank of Canada
SETI series titles	Short term loans
SOPH source publication	BANK OF CANADA REVIEW
WORD list of all terms in text fields	consumer

Indicate below what you wish to browse by typing the name of the type of information you have, e.g. SOPH.

Use the FIND command to perform a direct search.

TYPE OF BROWSE: BROWSE WORD FISH

Screen No. 10
CANSIM BROWSE KEYWORD INDEX

CANSIM / Search: Browse WORD FISH
Result filed under the following headings:

- 3) Word: FIRMS (690 series)
- 2) Word: FIRST (4328 series)
- 1) Word: FISCAL (129 series)
- 0) Word: FISH (332 series)
- 1) Word: FISHERY (1 series)
- 2) Word: FISHING (3688 series)
- 3) Word: FITTING (2 series)
- 4) Word: FITTINGS (7 series)
- 5) Word: FIVE (926 series)
- 6) Word: FIXED (1053 series)
- 7) Word: FIXTURE (63 series)
- 8) Word: FIXTURES (103 series)

Headings continue; press RETURN to see next page

To see a heading's series: type DISPLAY followed by a heading number.
To see a full series: type DISPLAY FULL followed by a number.
To begin a new search: type FIND or BROWSE.
To get more information: type HELP or OPTIONS.
YOUR RESPONSE: DIS 1
