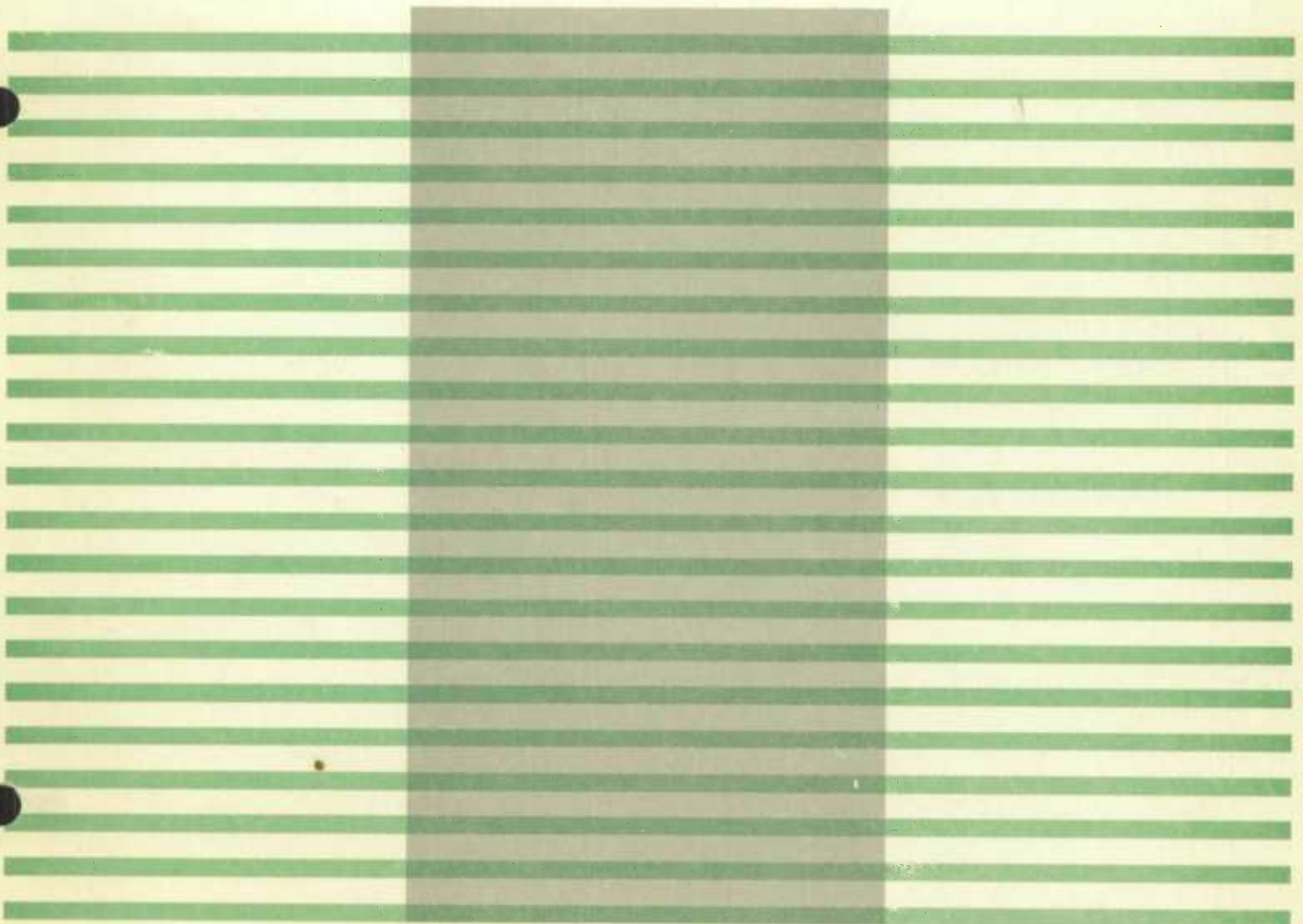


CANSIM:

USERS' MANUAL FOR DATA RETRIEVAL AND MANIPULATION



DOMINION BUREAU OF STATISTICS

The contents of this document may be used freely but DBS should be credited when republishing all or any part of it.



CANADIAN SOCIO-ECONOMIC INFORMATION
MANAGEMENT SYSTEM (CANSIM)

SERIES DIRECTORY

DOMINION BUREAU OF STATISTICS

MANAGEMENT SCIENCE

THE UNIVERSITY OF CHICAGO

PHOTOGRAPHED BY THE UNIVERSITY OF CHICAGO

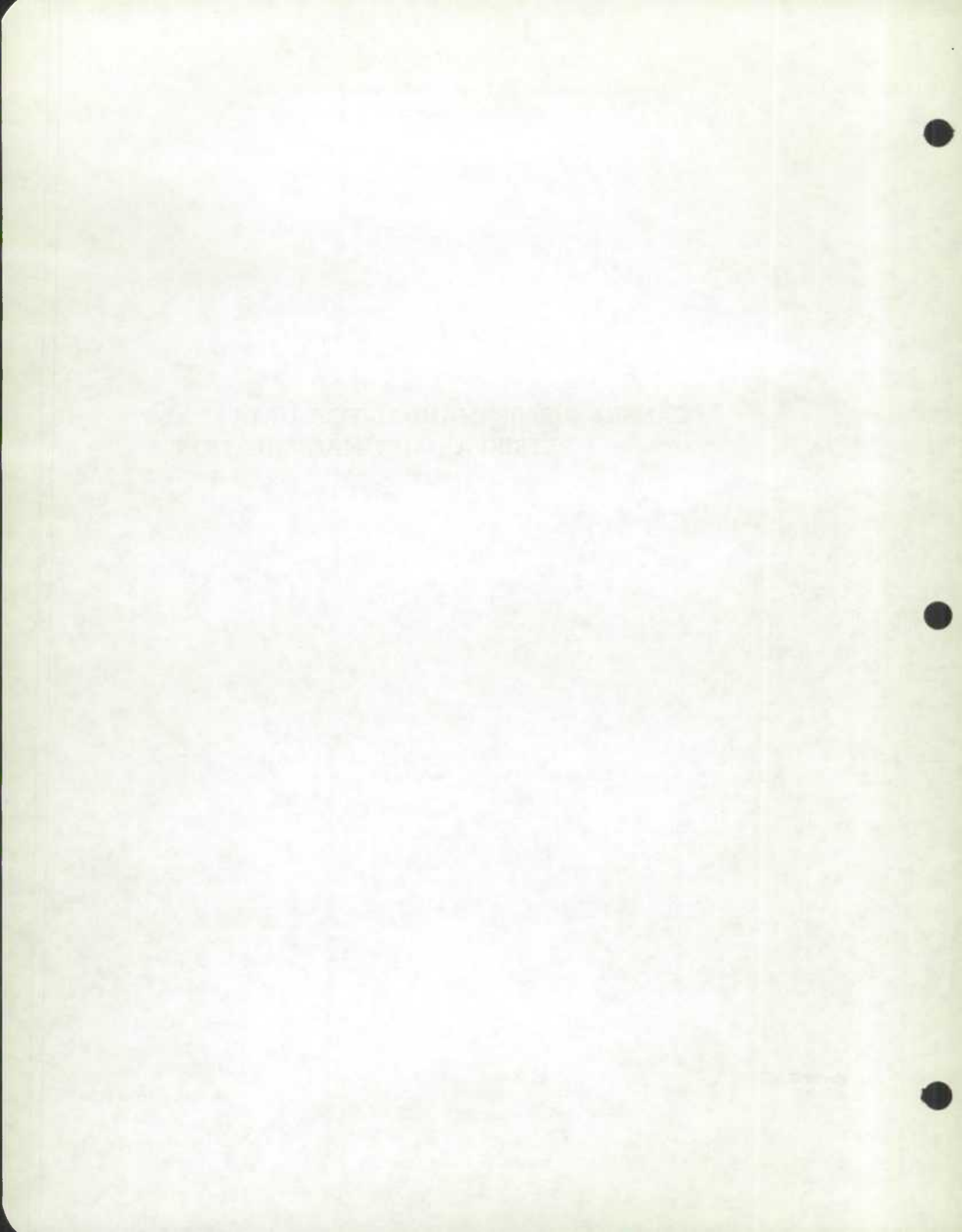
DOMINION BUREAU OF STATISTICS
National Accounts, Production and Productivity Division
General Time Series Section

CANSIM: USERS' MANUAL FOR DATA
RETRIEVAL AND MANIPULATION

Published by Authority of
The Minister of Industry, Trade and Commerce

May 1969
2204-505

Price: \$1.50



PROLOGUE

This manual describes part of a system which had its inception in a data storage, retrieval and manipulation computer package developed by M.C. McCracken. This prototype system was developed in 1964 at Southern Methodist University, where there was a need to collect and manipulate time series data in order to estimate parameters for an econometric model. The first version of the system used card images stored on magnetic tape and a small retrieval program which simply reformatted the data for input to statistical utility programs. In January 1965 the development of a more advanced system was started and a working version of this new system was in use by April of 1965.

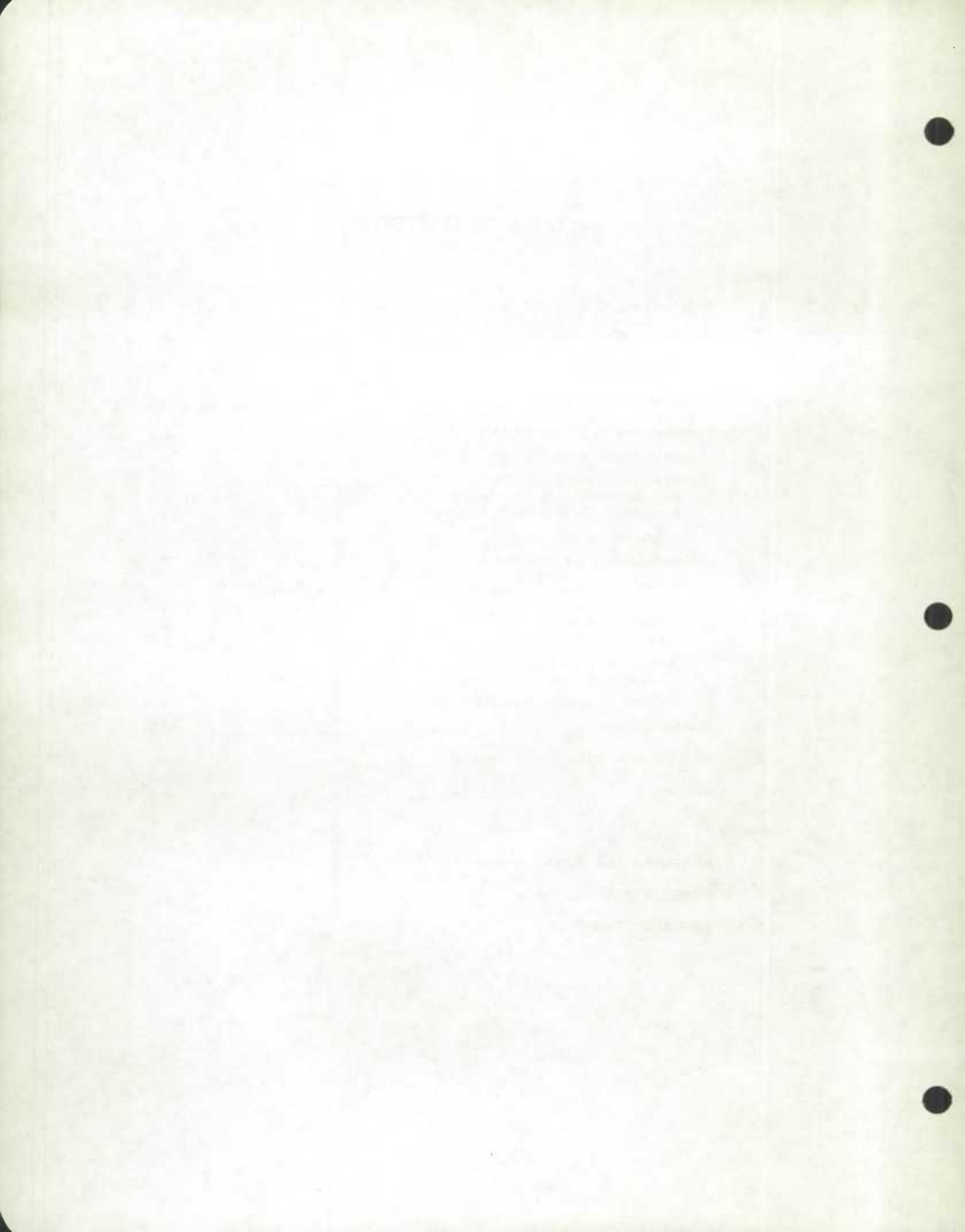
The Economic Council of Canada provided funds for the development of an expanded system on a CDC 3400 computer at the University of Montreal. The expanded version has been in use, with modifications, since September 1965. In May 1966 the Bank of Canada became the first agency other than the Council to make use of the system and during the Summer and Fall of 1966 the National Energy Board and the Department of Finance also began using the system for maintenance and manipulation of the data necessary in their analytical operations.

In November of 1966 the Dominion Bureau of Statistics accepted the responsibility for the entry of data into the base and maintenance of the existing programs. The Economic Council and the Bank of Canada expressed the hope that this system would eventually be modified into a true information system for use in the operations of statistical agencies of the Canadian government.

As a result, in July 1967, an inter-departmental team was set up under the direction of Dr. T.J. Vander Noot to design and implement a national data base for socio-economic data. This manual comprises one volume of the documentation for this system.

TABLE OF CONTENTS

Section		Page
1.	Introduction	1.1
2.	Description of the Data Base	2.1
	Matrix and Series Structure	2.1
	Matrix and Series Directory	2.2
3.	Retrieval and Manipulation Language	3.1
	General Description	3.1
	Requests for Retrievals	3.1
	Outside the Federal Government	3.1
	Agencies of the Federal Government	3.1
	Job Request Cards	3.1
	Conventions in the Command Set	3.3
	Commands	3.3
	Retrieve on Tape Commands	3.4
	Massager Format	3.4
	Publication Format	3.4
	Retrieve on Card Commands	3.5
	In Re-entry Format	3.5
	Retrieve in Table Format Command	3.5
	Error Messages	3.8
4.	Administrative and Billing Procedures	4.1
5.	Glossary	5.1
Appendix		
1.	MASSAGER Tape Format	
2.	PUBLICATION Tape Format	
3.	RE-ENTRY Tape Format	



INTRODUCTION

CANSIM (Canadian Socio-Economic Information Management System) is designed to provide efficient and economic management of a large volume of time-series data. The programs for data storage, retrieval, and manipulation comprising the system were written for the IBM 360/65. Management, control, and maintenance of the system are the responsibility of the Dominion Bureau of Statistics but accuracy of the included data is the responsibility of the agency compiling it.

Operation of the programs will be supervised by Data Bank Control, in the Operations Unit of the General Time Series (GTS) Section.

The subject of this manual is the retrieval sub-system which provides for the retrieval and manipulation of data stored in the base. A companion manual, entitled "CANSIM: Operation Manual for Data Entry" (Catalogue No. 12-530 Occasional—\$1.00), deals with the clerical and machine procedures used for entry, up-date, and revision of the data, is available from Publication Distribution.

The following sections describe the data base, the data directory, and the command language. While it is expected that the system will eventually be a real-time system, (providing immediate response to retrieval or manipulation requests by users) a batch processing mode will be followed at first.

Retrieval requests will be submitted to the CANSIM Operations Unit for batching with other such requests, for action at the next earliest running of the system.

A description of the data base (including record formats and explanations of codes) is given in Section 2. Also included are descriptions of the matrix and series numbering system and examples in the form of the Matrix Directory.

Section 3 is a description of the command language used to retrieve and manipulate data. Error messages are listed under a separate heading in this section. Included in this section is an outline of the job control cards necessary for correct delivery and billing.

Section 4 contains an outline of the administrative and billing procedures to be followed when submitting jobs.

The final section, Section 5, is a glossary of all words used in the command language or in the control or job cards. Samples of typical requests are also shown in this section.

Since the retrieval sub-system is considered to be evolutionary in nature, this entire manual is in loose-leaf form. As new commands are added to the sub-system or other changes are made, the appropriate pages will be revised and reissued.

DESCRIPTION OF THE DATA BASE

Each file or most probably time series in the CANSIM base is entered as part of a matrix of similar files arranged in hierarchical fashion. An

illustration might be a population table arranged as follows:

February, 1969

Population statistics

Table 1: population, by province (thousands)

years and months	Canada	Nfld.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Yukon	N.W.T.
1966 June	20,015	493	109	756	617	5,781	6,961	963	955	1,463	1,874	14	29
1967 June	20,405	500	109	757	620	5,868	7,149	963	958	1,490	1,947	15	29
1966 Jan.	19,857	490	108	754	616	5,740	6,888	962	952	1,456	1,848	15	28
Apr.	19,939	492	108	755	616	5,762	6,926	963	954	1,459	1,862	14	28
June	20,015	493	109	756	617	5,781	6,961	963	955	1,463	1,874	14	29
July	20,050	494	109	756	617	5,788	6,979	963	956	1,465	1,880	14	29
Oct.	20,158	496	109	755	617	5,812	7,033	961	957	1,470	1,905	14	29
1967 Jan.	20,252	497	109	755	618	5,833	7,078	959	956	1,476	1,927	15	29
Apr.	20,334	500	109	756	619	5,854	7,115	961	955	1,483	1,938	15	29
June	20,405	500	109	757	620	5,868	7,149	963	958	1,490	1,947	15	29
July	20,441	501	109	758	621	5,873	7,167	965	958	1,493	1,952	15	29
Oct.	20,548	502	109	758	623	5,894	7,217	966	959	1,502	1,973	15	30
1968 Jan.	20,630	502	110	760	623	5,910	7,252	968	959	1,511	1,990	15	30
Apr.	20,700	505	110	760	624	5,923	7,283	969	959	1,520	2,002	15	30
June	20,744	507	110	760	624	5,927	7,306	971	960	1,526	2,007	15	31
July	20,772	508	110	760	625	5,930	7,321	972	961	1,529	2,010	15	31
Oct.	20,857	511	110	762	626	5,945	7,355	974	962	1,538	2,028	15	31

Source: Estimated population of Canada, by province (91-201), D. B. S.

This table appears monthly in the Canadian Statistical Review. In the CANSIM data base, the time series (columns of data) have been restructured:

- 01 Total Canada
- 02 Newfoundland
- 02 Prince Edward Island
- 02 Nova Scotia

The entire "Table" is called a matrix. The "01" level within the matrix signifies that this time series is the total or summary measure. The "02" levels are thus subordinate in some way. Since

data collected as a single time series are almost always interdependent with other data, the matrix arrangement allows a whole set of files to be updated or revised at the same time. Matrices also allow for a greater degree of internal verification of the data entered. For instance, in the above example, the "02" level entries (Provinces) must add to the "01" total level (Canada).

All retrievals are made by a single number which indicates the matrix and series desired. The numbering scheme is illustrated below in the sample of the first page of the Matrix and Series Directory. This particular table is identified as Matrix 1.

MATRIX AND SERIES DIRECTORY

000001 NUMBER OF PERSONS IN CANADA, BY PROVINCES, BY QUARTERS SINCE 1940.

ESTIMATED POPULATION OF CANADA BY PROVINCE (91 -201), DBS.

ESTIMATES FOR CALENDAR QUARTERLY PERIODS, FROM JULY 1, 1951. QUARTERLY DATA RELATE TO JAN. 1, APR. 1, JULY 1 AND OCT. 1.

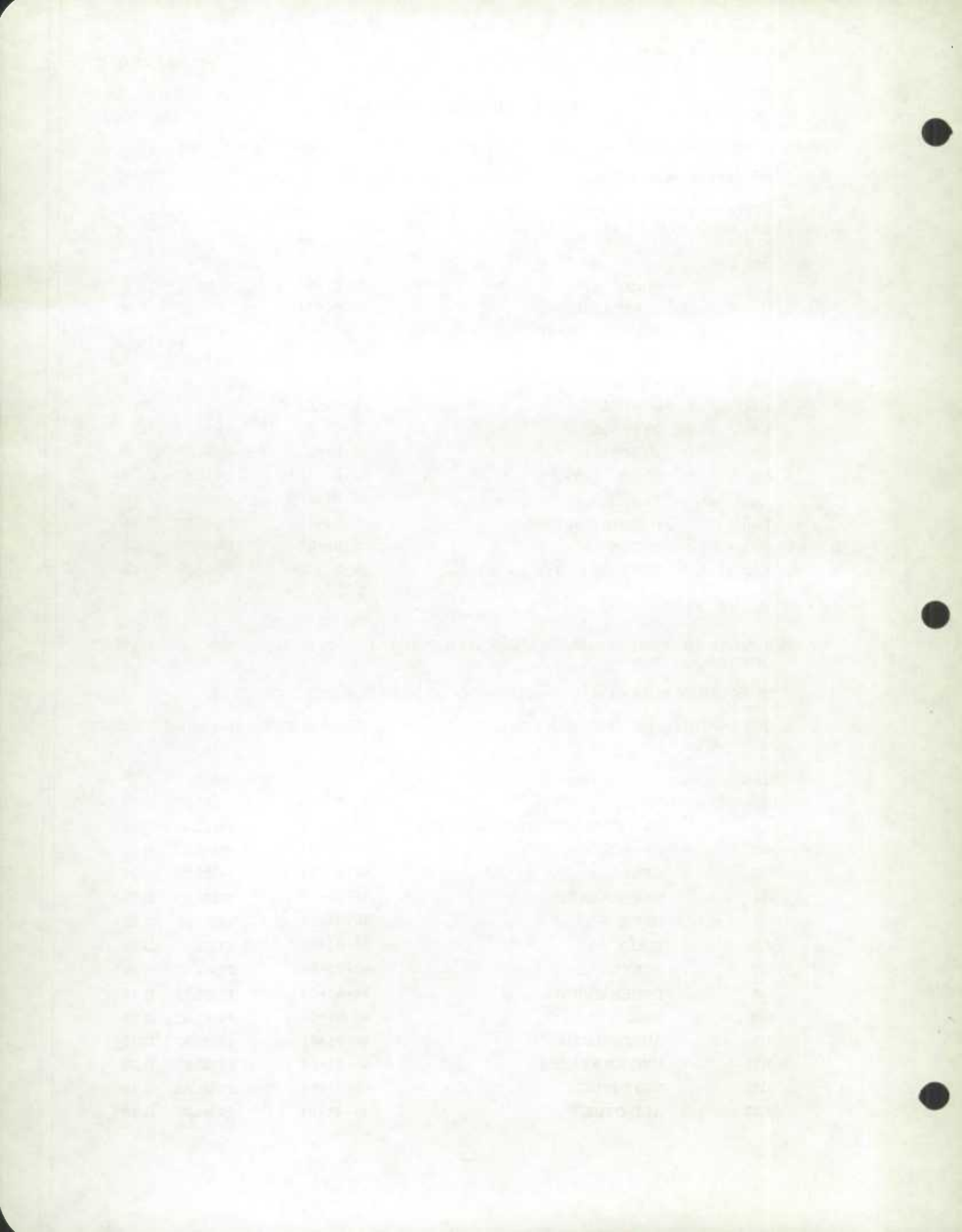
DBS 2	6002			
1	CANADA	40-01-01	PUBLIC	D 1
1.1	NEWFOUNDLAND	46-01-01	PUBLIC	D 2
1.2	PRINCE EDWARD ISLAND	47-01-01	PUBLIC	D 3
1.3	NOVA SCOTIA	40-01-01	PUBLIC	D 4
1.4	NEW BRUNSWICK	40-01-01	PUBLIC	D 5
1.5	QUEBEC	40-01-01	PUBLIC	D 6
1.6	ONTARIO	40-01-01	PUBLIC	D 7
1.7	MANITOBA	40-01-01	PUBLIC	D 8
1.8	SASKATCHEWAN	40-01-01	PUBLIC	D 9
1.9	ALBERTA	40-01-01	PUBLIC	D 10
1.10	BRITISH COLUMBIA	40-01-01	PUBLIC	D 11
1.11	YUKON	40-01-01	PUBLIC	D 12
1.12	NORTHWEST TERRITORIES	40-01-01	PUBLIC	D 13

000002 NUMBER OF PERSONS IMMIGRATING TO CANADA, BY COUNTRY OF LAST PERMANENT RESIDENCE.

QUARTERLY BULLETIN, DEPARTMENT OF MANPOWER AND IMMIGRATION.

DATA PUBLISHED APPROXIMATELY 75 CALENDAR DAYS AFTER END OF REFERENCE QUARTER.

DBS 2	6002			
1	TOTAL	46-01-01	PUBLIC	D 27
1.1	UNITED KINGDOM AND IRELAND	46-01-01	PUBLIC	D 28
1.2	FRANCE	46-01-01	PUBLIC	D 29
1.3	GERMANY	46-01-01	PUBLIC	D 30
1.4	NETHERLANDS	46-01-01	PUBLIC	D 31
1.5	GREECE	46-01-01	PUBLIC	D 32
1.6	ITALY	46-01-01	PUBLIC	D 33
1.7	PORTUGAL	46-01-01	PUBLIC	D 34
1.8	OTHER EUROPE	46-01-01	PUBLIC	D 35
1.9	ASIA	46-01-01	PUBLIC	D 36
1.10	AUSTRALASIA	46-01-01	PUBLIC	D 37
1.11	UNITED STATES	46-01-01	PUBLIC	D 38
1.12	WEST INDIES	46-01-01	PUBLIC	D 39
1.13	ALL OTHER	46-01-01	PUBLIC	D 40



These directory entries show the basic structure of the base. Everything needed by the user on the "matrix record" is shown in the directory. First is shown the matrix number which identifies the specific matrix in all retrievals. Next are shown the long title and the source of the data. Following is the matrix note (if any) and finally the agency and section responsible for the data. All of the above material is part of the matrix record and pertains to all subsidiary data series.

The data series are listed next. Each line shows the series number, the title, the date of the earliest data, and finally whether or not there are any restrictions on the use of the data. The term "PUBLIC" means that there are no restrictions,

"PART SEC" indicates that at least one data point is restricted, and "SECURE" indicates that the entire series is restricted. The last entry is the Data Bank series number. Further, if a series is terminated, "T" will appear to the left of the series-number and data may be retrieved from a terminated series.

The directory is considered to be a full statement of the contents of the data base and updates and revisions will be made as required.

A full and detailed statement of the contents of the matrix and series records is shown in Appendix 2 as part of the documentation of the "Retrieve in Publication Format" command.

RETRIEVAL AND MANIPULATION LANGUAGE

The retrieval and manipulation language is designed to be as flexible and powerful as is possible, while at the same time staying within the staff and time restrictions which exist for programming and analysis.

One problem that has existed with some command sets is that they are not designed to be added to, and if additional commands were necessary, the entire structure of the language had to be revised. It is a goal of the CANSIM system that additional commands may be added without altering previously existing commands.

Another design criterion was that the users of the system will be primarily subject-area specialists with little or no interest in how the "big black box" actually performs the tasks given to it. The manipulation sub-routines are controlled by unformatted English language commands (with user-oriented syntax and vocabulary) and this manual does not explain the program's logic which interprets the commands.

Experienced programmers tend to prefer command languages that are terse and have little redundancy. Such compact languages, however, tend to be difficult for non-programmers to learn and use; therefore, the command set tries to partially satisfy both the programmer and non-programmer by having a long and short form for each command.

Requests for Retrievals

In the handling of information requests a distinction is made between Federal Government and all other users. This difference is made necessary since government users have access to the Central Data Processing Service Bureau.

All keypunch instructions below assume the EBCDIC code rather than BCD.

Retrieval Requests Originating Outside the Federal Government

Non-government users should refer to Section 4 Part B for procedures to follow for retrieval requests. They should however read the next section which deals with the procedures for government users. Retrieval options available to government users are also available to non-government users.

Requests Originated by Agencies of the Federal Government

Retrieval requests and data entry cards are to be sent directly to the CANSIM Clerk, General Time Series Section, DBS. Request cards must be completely keypunched according to the formats shown below.

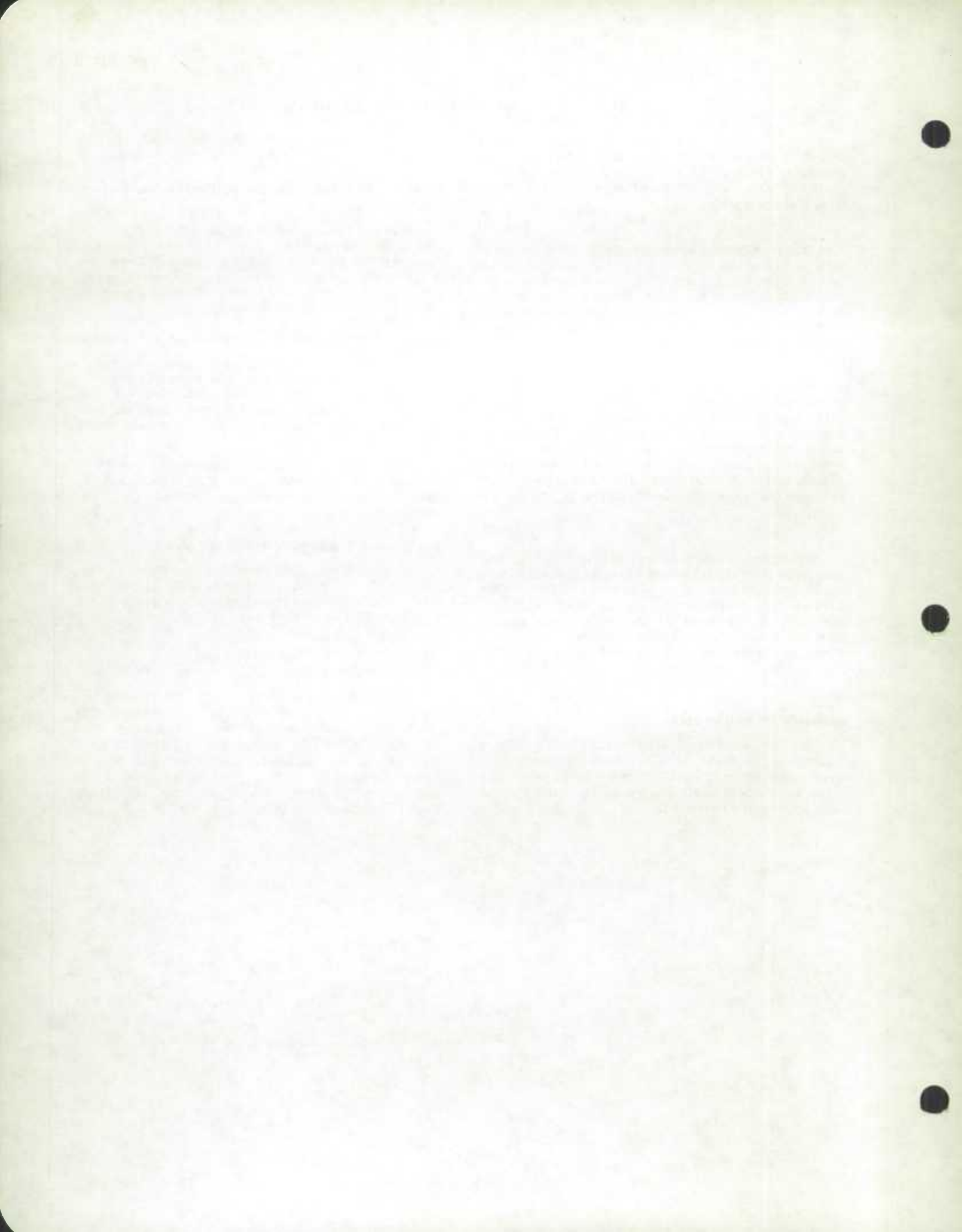
Retrievals and data entry printouts will be returned to the originator by the CANSIM Clerk, or in some instance arrangements may be made for direct delivery from CDPSB.

Job Request Cards (Government Users)

Requests received must be identified so that the job can be returned and the cost of the work billed appropriately. Therefore, each request or job consists of three or more cards:

1. A START JOB card;
2. One or more command cards; and
3. A FINISH JOB card.

Many jobs (requests from different people) will be batched together to be run on the computer. The jobs, then, are separated one from another by a FINISH JOB and the next job's START JOB card. All jobs being batched together as one run are independent of each other; that is, an error in one job causes the termination of only that job and not the other jobs in that batch.



The format and Hollerith Code for START JOB card is given below:

Column number	Field type ¹	Explanation
1 - 4	A	RSFC. An identifying mnemonic for CANSIM retrieval cards.
5 - 8	AN	A four letter mnemonic identifying the agency. A full list of these codes is available from Data Bank Control.
9-11	AN	Any identification acceptable to the user agency adequate to identify the specific user. This may be a persons initials or some "internal" agency accounting code.
12-15	N	Card Number. "For safety's sake" all cards in a job should be sequentially numbered.
16-25	A	Punch 'STARTbJOBb'.
26-80	AN	Heading. The contents of this field will appear at the top of each page of printed output or be attached to any tape reel or card output from the request.

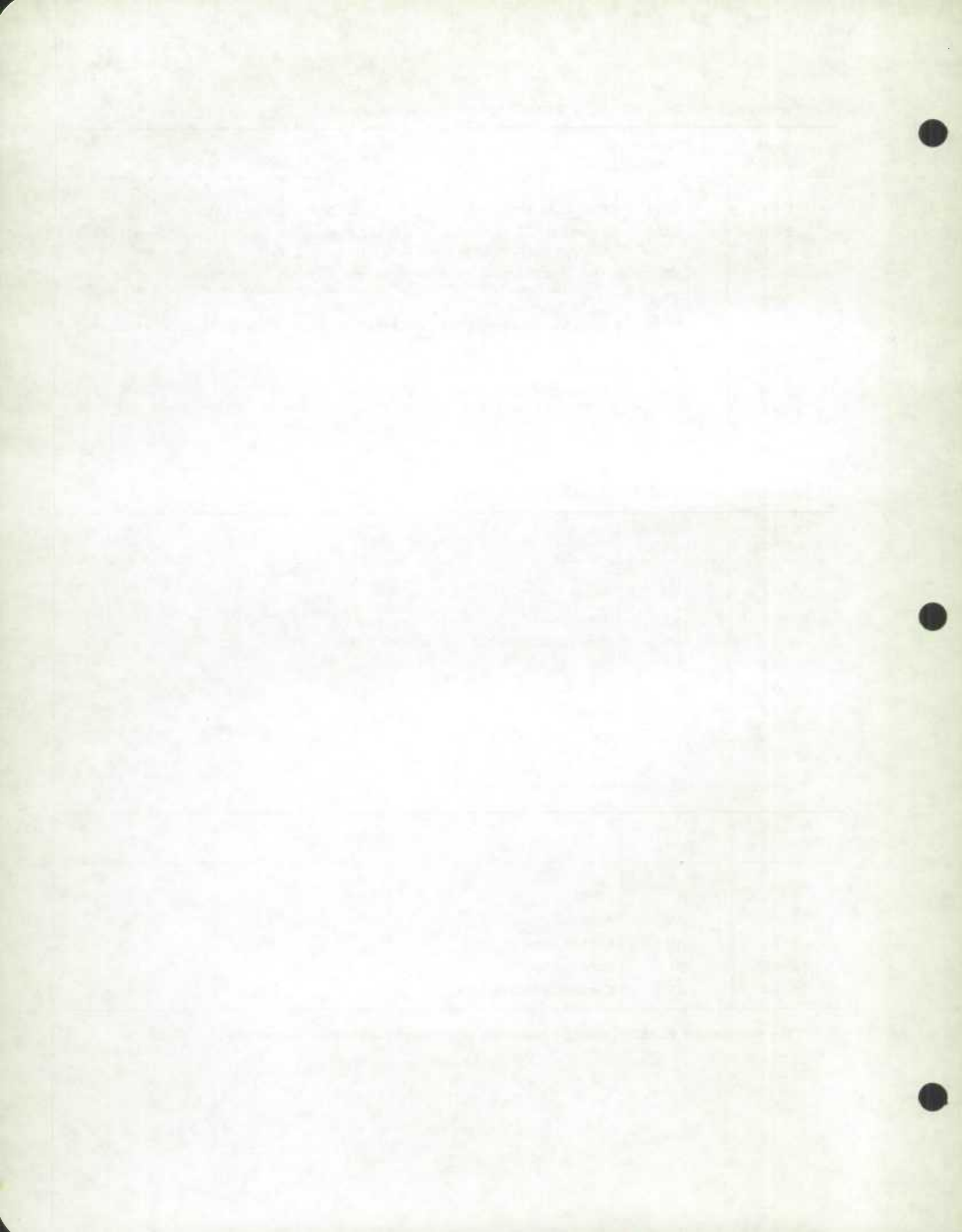
The format of the FINISH JOB card is as follows:

Column number	Field type ¹	Explanation
1 - 4	A	RSFC
5 - 8	AN	Agency name
9-11	AN	User in agency code (as in START JOB card).
12-15	N	Card number
16-25	A	'FINISHbJOB'
26-80		Blank

The format of the Command Cards is similar:

Column number	Field type ¹	Explanation
1 - 4	A	RSFC
5 - 8	AN	Agency
9-11	AN	User in agency
12-15	N	Card number
16-80	AN	Commands in free form

¹ The abbreviations A, AN, N, stand for alphabetic, alphanumeric, and numeric respectively.



Conventions in the Command Set

1. Each command must end with a period. No other punctuation is necessary.
2. A four digit card number in a job is optional. If it is used, a card sequence check will be done. Out of sequence card(s) will result in warning messages only, not a job termination.
3. In the event that a command is continued on several cards, put only **complete** words on each card. If one word ends in column 80, then the next card must have a blank in column 16.
4. Upper case words are CANSIM words and must be spelled as they appear.
5. Upper case words which are underlined are words making up the abbreviated command set. These are called "key-words."
6. Upper case words which are not underlined are optional and may be omitted.
7. Lower case words describe information which must be supplied by the user.
8. Square brackets ([]) indicate optional commands.
9. Braces ({ }) enclosing a list means that the programmer **must** select one of the enclosed items.
10. A word or phrase enclosed by a pair of slashes (//) indicates that the word or phrase may be repeated a number of times.

Retrieve Command

RETRIEVE { series-identifier }
LIST

This command seeks one or more series records in the data-base. A specific series-identifier or the word LIST must always follow the word RETRIEVE.

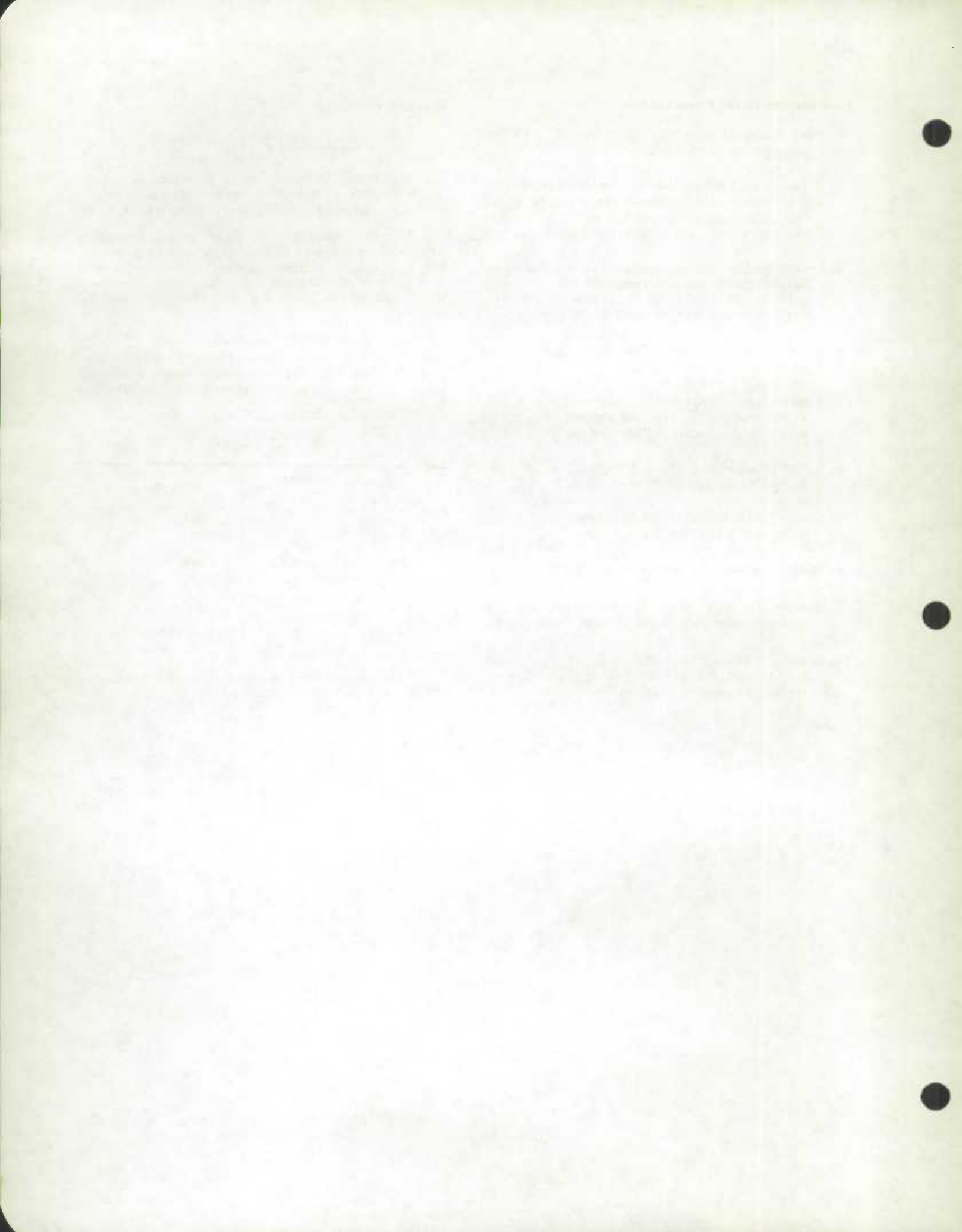
A series identifier is made up of the matrix number and the series number separated by a period. For instance, to retrieve the series on immigration from Greece, the number 2.1.5 would be used (see Matrix and Series Directory). Leading zeros may be omitted.

When the word LIST is used, it indicates that a number of series are to be retrieved and series-identifiers will appear on a set of cards (one card per series) following the command card as follows:

RETRIEVE LIST
//series-identifier//

Column number	Field type ¹	Explanation
1- 4	A	RSFC
5- 8	AN	Agency
9-11	AN	User in agency
12-15	N	Card number
16-80	AN	Series number

¹ The abbreviations A, AN, N, stand for alphabetic, alphanumeric, and numeric respectively.



Retrieve on Tape Commands

RETRIEVE { series-identifier }
LIST } ON TAPE IN MASSAGER FORMAT.

The basic function of the command is that it seeks the series records in the data base, changes their formats, and writes the retrieved and reformatted series out on tape. The data will be converted to floating point numbers (single precision) and the tape may be used as input to DATABANK, MASSAGER, and MATOP programs running on IBM 360's. Tape conversions in most cases will be needed for using

it as input to other computers. A FORTRAN program to convert to 7 track BCD tape is available at no cost. If double precision is required the words "IN DOUBLE PRECISION" must follow the word "FORMAT" i.e. FORMAT IN DOUBLE PRECISION.

See Appendix 1 for the tape layout.

RETRIEVE { series-identifier }
LIST } ON TAPE IN PUBLICATION FORMAT.

Same as for MASSAGER format except that the PUBLICATION format contains all of the information stored in the base pertaining to that series and is

used primarily as input to report generating programs for printed publications. (See Appendix 2 for the tape layout.)

RETRIEVE { series-identifier } ON TAPE IN { MASSAGER
LIST } { PUBLICATION } FORMAT

[USING SECURITY security-word].

In order to retrieve series which have security restrictions, the correct seven character (alpha-numeric) word must be used following the key-word SECURITY. Only the non restricted data of a partially restricted series will be retrieved if the proper SECURITY word is missing. Any errors in spelling will block retrieval. Notice is always sent to the responsible agency of any retrieval or attempted

retrieval of a restricted series using the SECURITY option.

Retrievals of secured data are somewhat more complicated when the LIST option is used since the necessary security word may be different for each series retrieved. If the word as shown is used in the security option then the format of the series identifier card also changes. For instance:

RETRIEVE LIST ON TAPE IN { MASSAGER
LIST } { PUBLICATION } FORMAT USING SECURITY AS SHOWN.

//series-identifier security-word//

If public series, needing no code word, are included in the list of series to be retrieved then the security-word may be omitted for those series.

RETRIEVE { series-identifier }
LIST } ON TAPE IN MASSAGER FORMAT

[USING SECURITY security-word]

[GIVING NEW-SECURITY new-security-word].

MASSAGER format tapes can be used as work tapes and it sometimes is worthwhile to restrict data. The use of NEW-SECURITY adds an eight character field to the MASSAGER format tape which makes it a restricted series under the MASSAGER program.

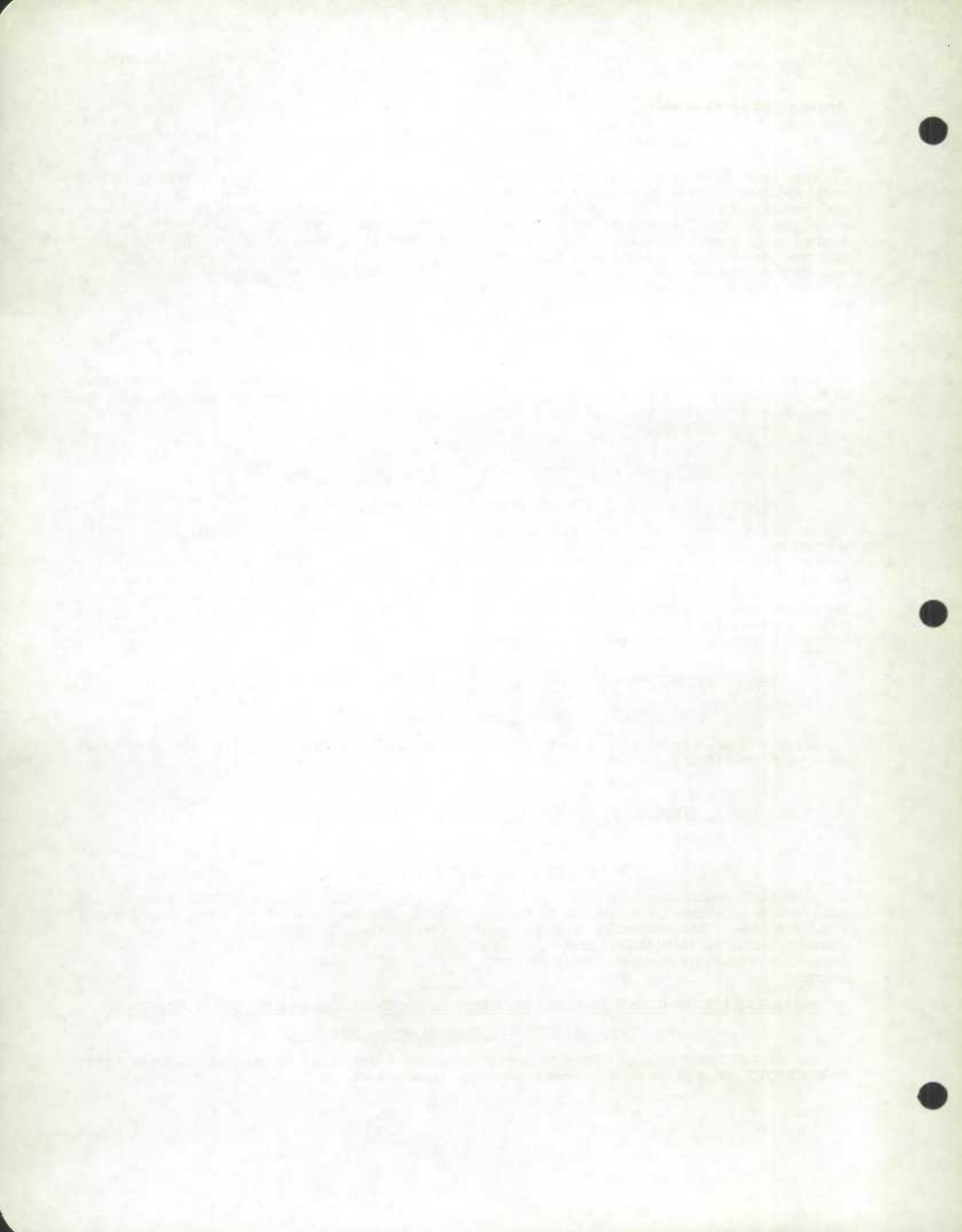
If the LIST, and NEW-SECURITY options are used, then the format of the series identifier card changes again. To illustrate:

RETRIEVE LIST IN MASSAGER USING SECURITY AS SHOWN GIVING NEW-SECURITY AS SHOWN.

//series-identifier security-word new-security-word//

The NEW-SECURITY option cannot be used if the SECURITY option is not used. In other words, a

public series cannot be retrieved and given a new-security-word.



Retrieve on Card Commands

RETRIEVE { series-identifier } IN RE-ENTRY FORMAT
LIST

[USING SECURITY security-word].

The output tape in the RE-ENTRY format contains card images. This tape can be re-entered into the base through the data-entry program of the CANSIM system. However, in order to re-enter the

data into the base, the data entry and other security words, if necessary, must be added to the cards since the retrieval program will **not** retrieve security words and display them in any way. (See Appendix 3).

Retrieve in Table Format Command

This command produces a "working table" with which the user can examine the detail and

content of the data base. The general format of the command is:

RETRIEVE { series-identifier } IN TABLE FORMAT [USING SECURITY security-word].
LIST

The general format of the table is shown on the next two pages. There are a number of restrictions which are discussed below.

1. Provision is made for only ten series to be listed on one page. If the number of series-identifiers following the LIST option exceeds ten, the first ten series will be shown on the first table and the next ten on the next, etc.
2. The date column will be derived from the series with the greatest frequency for which a data value is available. For example, given a combination of quarterly and monthly series, the date column will be derived from the monthly series.
3. To retrieve a "SECURE" data point, the appropriate retrieval code-word is necessary. If a "SECURE" point is retrieved, an asterisk will be printed beside the number and a warning:

"THIS PAGE CONTAINS SECURE DATA" will be printed at the bottom of the page.

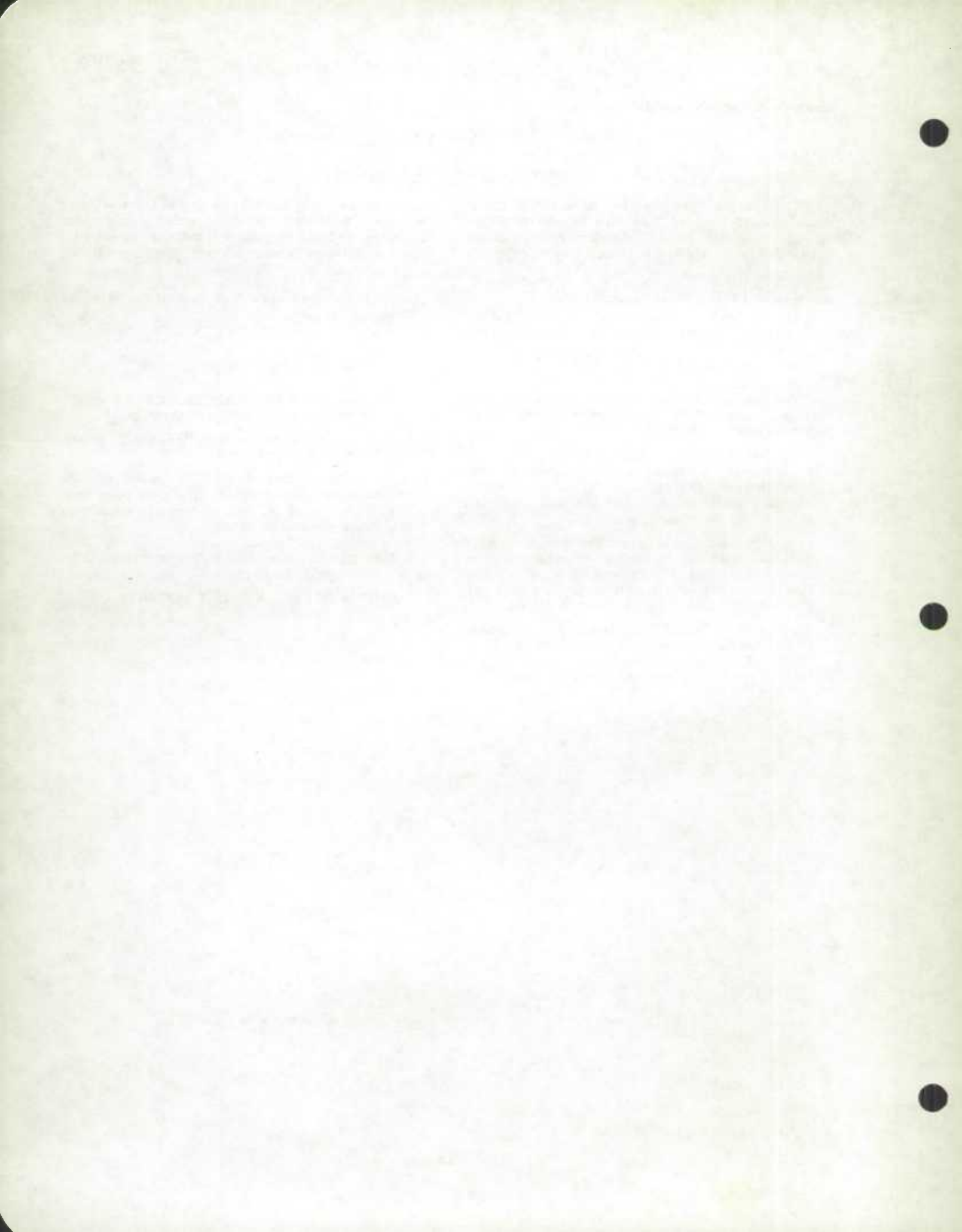
If no code-word is given, then "SECURE" points will be omitted or left blank.

4. There is no method to limit the amount of data printed at the present. If the first data point on a monthly series is in 1940, every data point from that time on would be listed.

The printout from the following command is shown on the next two pages:

RETRIEVE LIST IN TABLE FORMAT.

1.1
 1.1.1
 1.1.2
 1.1.3
 2.1



000001 NUMBER OF PERSONS IN CANADA, BY PROVINCES, BY QUARTERS SINCE 1940.

COLUMN 1: 1 CANADA

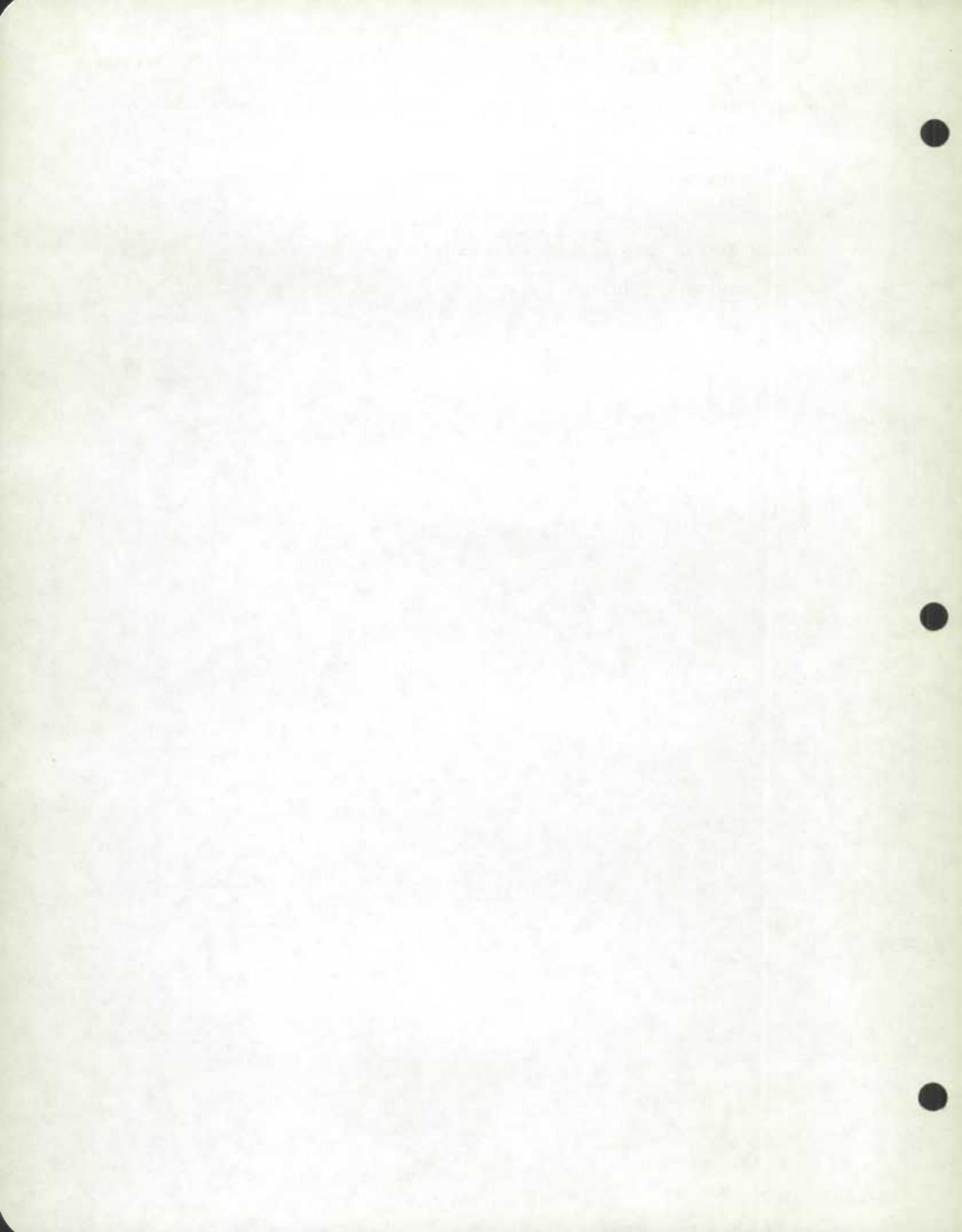
COLUMN 2: 1.1 NEWFOUNDLAND

COLUMN 3: 1.2 PRINCE EDWARD ISLAND

COLUMN 4: 1.3 NOVA SCOTIA

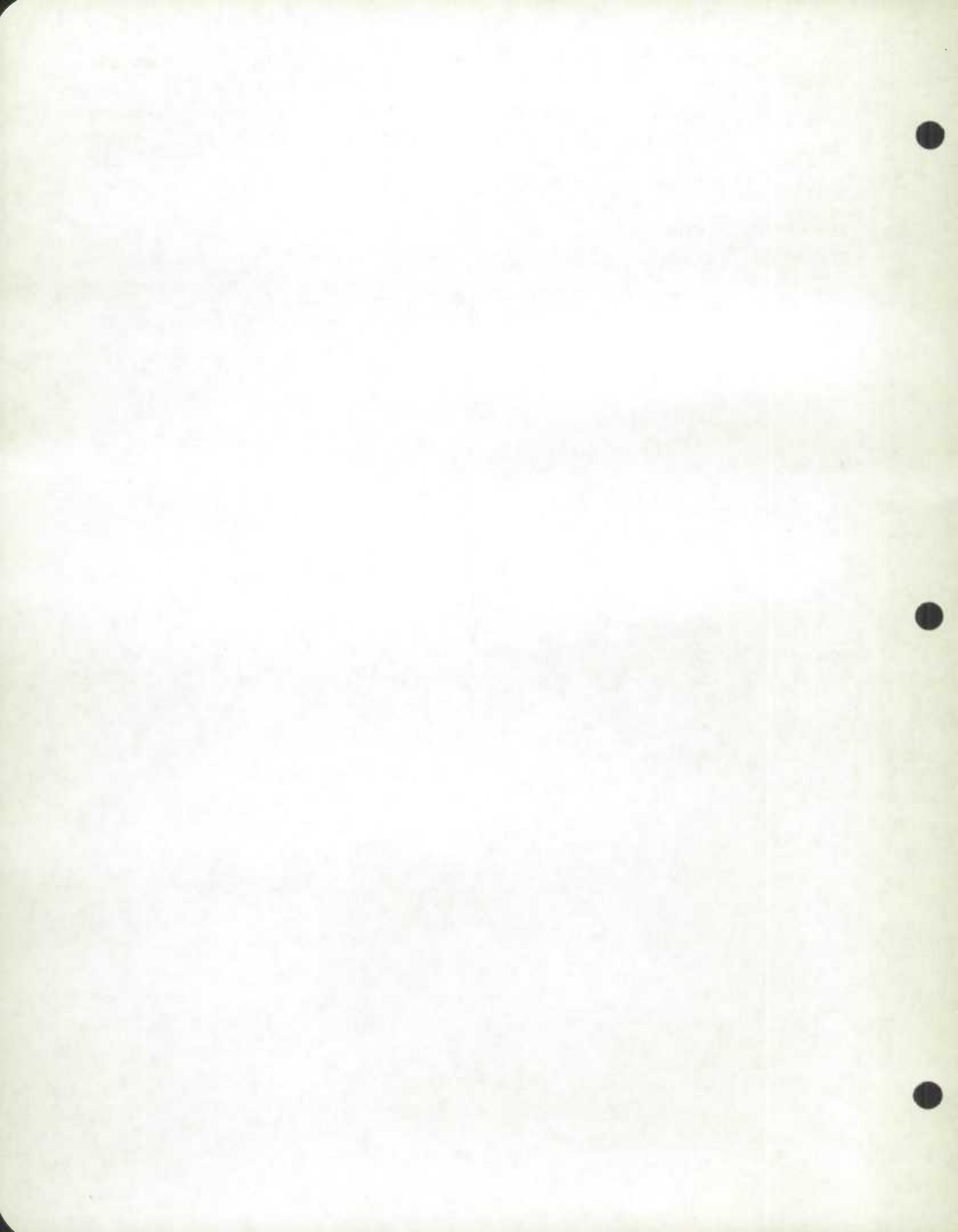
000002 NUMBER OF PERSONS IMMIGRATING TO CANADA, BY COUNTRY OF LAST PERMANENT RESIDENCE.

COLUMN 5: 1 TOTAL



SECTION 3

DATE	COLUMN 1 THOUSANDS OF PERSONS	COLUMN 2 THOUSANDS OF PERSONS	COLUMN 3 THOUSANDS OF PERSONS	COLUMN 4 THOUSANDS OF PERSONS	COLUMN 5 NUMBER OF PERSONS
JAN. 40	9,806	-	-	595	-
FEB. 40	9,819	-	-	595	-
			↓		
DEC. 45	11,622	-	-	620	-
JAN. 46	11,703	368	-	621	92,554
			↓		
JUN. 67	20,405	500	109	757	64,969
SEP. 67	20,552	501	109	758	72,803



Error Messages

The error messages shown below are in alphabetic order. A "T" (terminate) code means that the job is terminated at that point. An "I" (ignore) code indicates that **only** that card has been ignored. A "W" (warning) means that the program has made some assumption about the command and this assumption should be checked.

The error message appears immediately after the command to which it refers. Since the retrieval program is written as an "interpreter", each command is obeyed as it is read.

CARD IDENTIFICATION INCORRECT:

- I - If the card does not contain the identification mnemonic 'RSFC', this code is issued and the card is ignored. If this results in the START JOB or FINISH JOB card being missing, then that error message will show as well. If the START JOB card is missing, the job will be terminated and all cards will be ignored until the FINISH JOB or the next START JOB card is found.

CARD OUT OF ORDER:

- W - This message indicates that the card numbers (when used) are not in order.

CHANGE IN AGENCY CODE:

- I - If the agency code in subsequent cards is not the same as given on the START JOB card, the command will not be actioned.

COMMAND CANNOT BE INTERPRETED:

- I - If some essential element of the command is omitted such as the word RETRIEVE or the period, then this message will be given.

FINISH JOB CARD MISSING:

- W - This indicates that the FINISH JOB card is missing. If the next job is missing the START JOB card, all commands will be billed to the person and agency indicated on the first START JOB card if the agency remains the same.

IMPROPER FORMAT STATEMENT:

- I - A retrieve command requires that one format statement (MASSAGER, PUBLICATION, RE-ENTRY, or TABLE) be selected to indicate the desired output. If the word is mis-spelled or omitted, this message will appear beside the appropriate command.

IMPROPER SERIES-IDENTIFIER:

- I - The program expects the word after the key-word RETRIEVE to be either LIST or a series-identifier. If the word LIST is mis-spelled or the series-identifier has the wrong format, this error message will print out beside the command line on the print-out.

INCORRECT SECURITY-WORD:

- I - A valid security-word appropriate to the series being retrieved must be given in the command immediately following the word SECURITY.

INVALID AGENCY CODE:

- T - If the agency mnemonic given on the START JOB card is not in the list of acceptable codes, all commands in that job will be ignored.

NEW-SECURITY STATEMENT IS REDUNDANT:

- W - The NEW-SECURITY option has meaning only if applied to the MASSAGER format. If used in connection with any other format it is redundant and is ignored.

NEW-SECURITY-WORD TOO LONG:

- I - More than eight characters have been shown for a new-security-word.

REPORT FREQUENCY INCORRECT:

- I - The MASSAGER format permits only annual, quarterly, and monthly series to be retrieved. If an attempt is made to retrieve a series with another frequency, that card is ignored and this message printed out.

SECURITY-WORD TOO LONG:

- I - More than seven characters have been shown for a security-word.

SERIES IS NOT IN BASE:

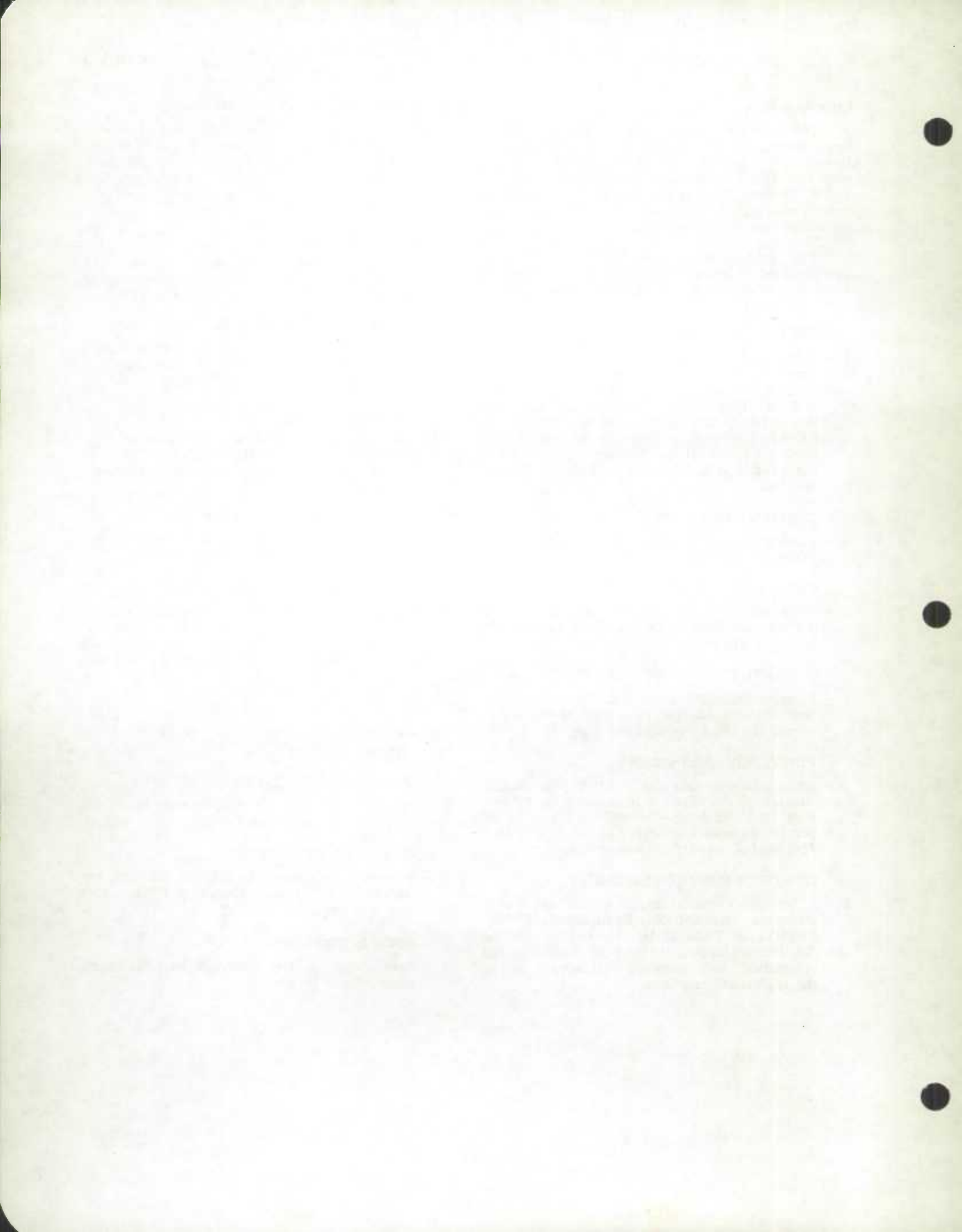
- I - This message indicates that the series requested cannot be located in the base.

START JOB CARD MISSING:

- T - This error terminates the job and all cards are ignored until the next START or FINISH JOB card is found.

WORD IS TOO LONG:

- I - Some word in the command exceeds thirty characters in length.



ADMINISTRATIVE AND BILLING PROCEDURES

This section is in two parts; Part A covers procedures to be followed by Federal Departments and Agencies who have an account with the Central Data Processing Service Bureau; and Part B covers procedures to be followed by non-government users.

PART A

All jobs (retrieval request cards and data entry cards) are to be submitted with a work ticket to the CANSIM Clerk, General Time Series Section, DBS. A rubber stamp with additional information required will be made available on request. The following information on the work ticket must be completed:

- (a) programmer's name
- (b) account number
- (c) phone number
- (d) date submitted
- (e) estimated running time
- (f) information on stamp

On receipt of the work ticket, the CANSIM Clerk will combine the work tickets from the various departments and agencies, and prepare one job submission to CDPSB under Dummy Account CSM01. A copy of all work tickets included in this job will be filed together with a copy of the job submission form.

Completed jobs will carry information provided by the CANSIM housekeeping system on **total** run time and times for **each** agency and section. On completion of each job the clerk will:

- (a) calculate the percentage of the job for which each agency and section is responsible for, and indicate this percentage on the work ticket; and
- (b) return completed runs to originator.

At the end of each month, CDPSB will provide the CANSIM Clerk with a statement listing the jobs submitted under Dummy Account CSM01 with cost for each job. This statement will not only include the time provided by the CANSIM housekeeping system but also the time and cost for additional operations such as SYSIN, SYSOUT, and SPOOL, which may have been incurred by the job.

For each job the clerk will calculate the cost to be borne by each work ticket and indicate the amount on the work ticket. For example a job costing \$25.00 with 3 work tickets will be calculated as follows:

Work Ticket 1	25% = \$ 6.25
Work Ticket 2	10% = \$ 2.50
Work Ticket 3	65% = \$16.25
	100% \$25.00

This means that the cost for additional operations for a given job is prorated. Although using this method may not be equitable for any given job, over a period of 3-4 months it should prove reasonable and fair.

After calculating the cost for each work ticket, the work tickets will be submitted to CDPSB for billing action. Under this procedure billing action may be one month in arrears.

PART B

Non-government users must submit all retrieval requests to the CANSIM Clerk, General Time Series Section, Dominion Bureau of Statistics, Ottawa.

A CANSIM Purchase Agreement should be submitted unless previous arrangements obviate this requirement. If the form does not provide sufficient space, a separate list should be attached indicating the type of retrieval with the series-identifier. Since command and series-cards are pre-keypunched by the General Time Series Section, cards will not be required.

The "standard tape" in MASSAGER format (single precision) which can be used as input to DATABANK and MASSAGER program is available only in 360/65 format. A FORTRAN program to convert to 7 track BCD tape is available at no cost. If double precision is required, additional cost will be incurred.

Requests for Purchase Agreement forms and inquiries should be directed to Mr. T. Tanaka, CANSIM Users' Service, General Time Series Section, DBS, Ottawa.

Faint, illegible text at the top of the page, possibly a header or introductory paragraph.

Second block of faint, illegible text in the middle of the page.

Third block of faint, illegible text in the lower middle section of the page.

Final block of faint, illegible text at the bottom of the page.

GLOSSARY

Command A group of words delimited by a period followed by a space on the input card which initiates the retrieval of a series in the specified format.

FORMAT:

Massager Tape generated of the requested series to be used as input to manipulative programs such as MASSAGER or MATOP.

Publication Tape generated of the requested series and used primarily as input to report generating programs for printed publications. It contains almost all of the information stored in the base pertaining to that series.

Re-entry A card-image tape of the requested series which can be re-entered into the base through the data-entry program of the CANSIM system.

Table This format produces a "working table" printout with which the user can examine the content and detail of the data base.

Key-word A word which initiates a specific action by the RSFC program.

List When used, the program will expect a set of cards with series-identifiers to follow.

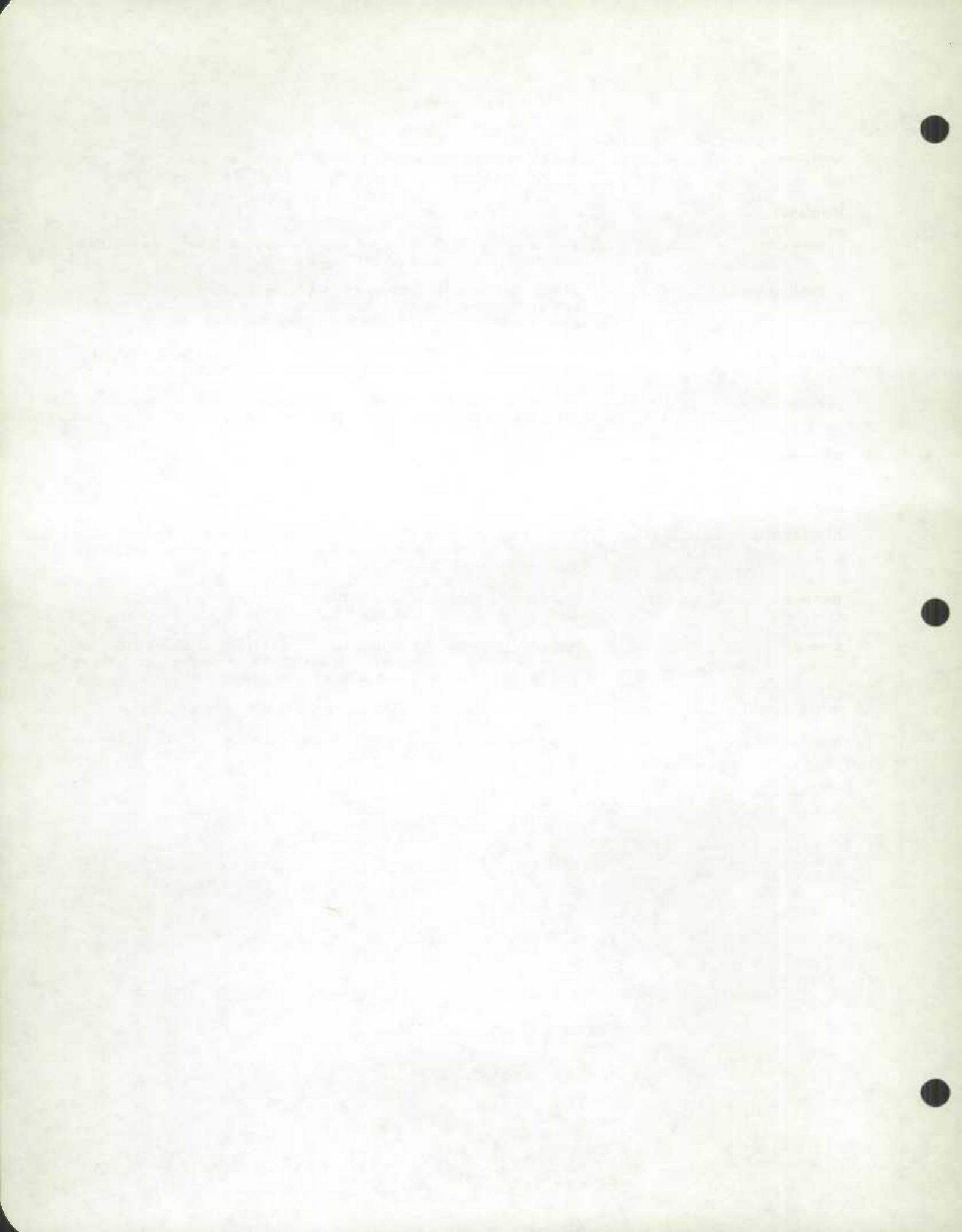
New-security The use of this word adds an eight character field to the MASSAGER format tape which makes it a restricted series under the MASSAGER program.

Retrieve Command that seeks one or more series records in the data-base. A series-identifier or the word LIST must immediately follow this key-word.

Security The RSFC program will interpret the seven character (alphanumeric) word following as the security word. If **SECURITY** is omitted, the program will generate a seven character blank security-word.

Series-identifier Made up of the matrix number and series separated by a period.

Word A group of consecutive non-blank characters in the command field of the input card. (Max. \leq 30 characters).



APPENDICES

Note: The formats following are necessarily brief. For full information it will be necessary to check the appropriate sections of the following publications: "DATABANK" and "CANSIM: Operation Manual for Data Entry".

MASSAGER COMPATIBLE BINARY TAPE FORMAT

CREATED BY CANSIM

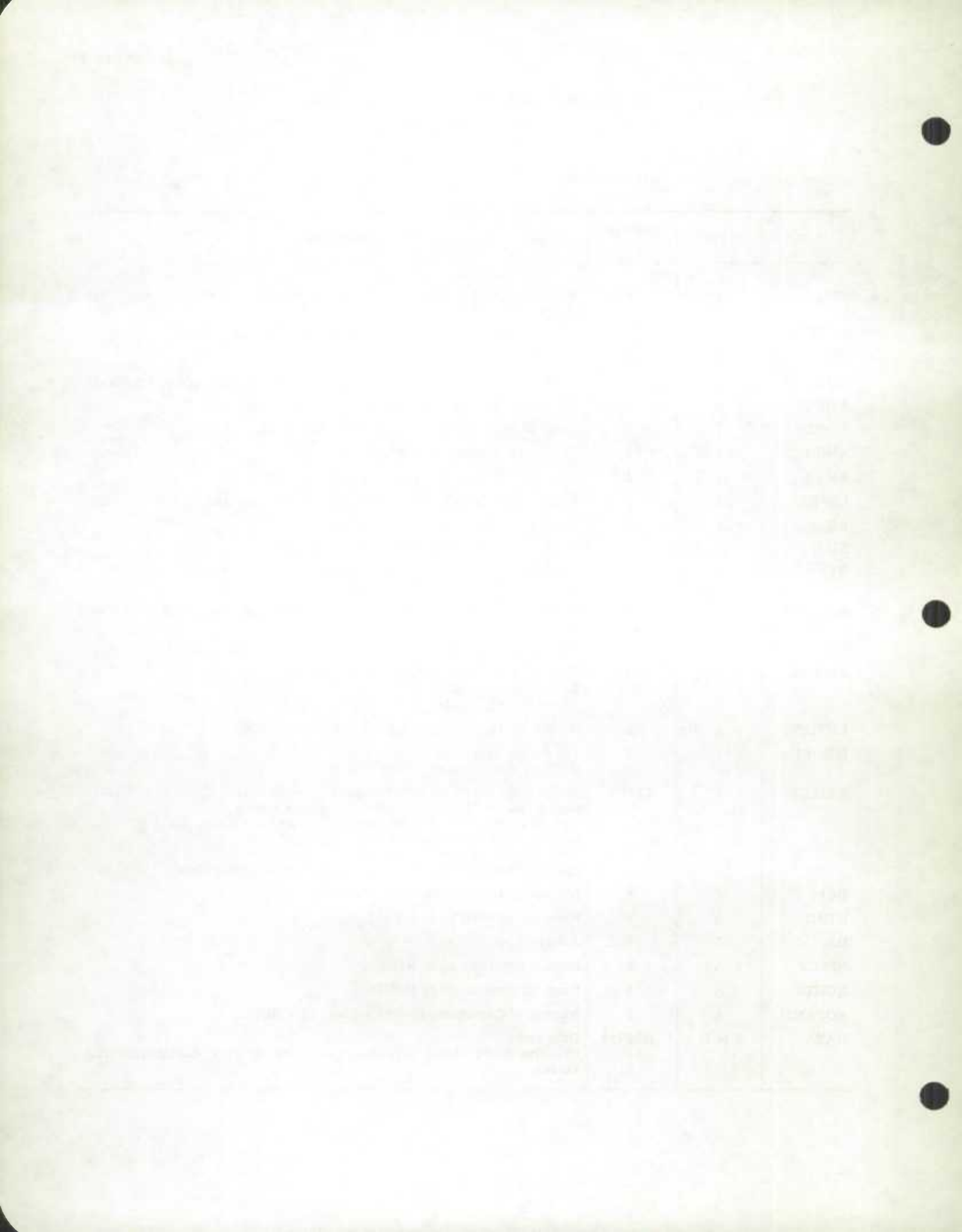
I = Integer

S = Single precision

A = Alphanumeric

D = Double precision

Name	Type	Number of bytes	Description
ST1	I	4	Total number of 8 byte words of title information in matrix, series, etc. (Always NOTIT x 10).
LABEL	A	8	An 8 character series identification code (DATABANK).
M1	I	4	Always 0.
NOSEC1	A	8	An 8 character series security code (series security code and a blank).
NOUT1	A	8	DB verification tag (last update in form YY-MM-DD).
NOPEN1	I	8	Length of title record in characters (ST1 x 8).
KBEG	I	4	First year of data series (e.g. 1928)
KEND	I	4	Last year of data series (e.g. 1968)
ESEC1	A	8	Blank. An 8 character edit security code used in DATABANK.
KIND	I	4	Annual = 1, Quarterly = 4, Monthly = 12.
NOTIT	I	4	Number of title cards. Equals N1 + N2 + N3.
N1	I	4	Number of series cards (i.e. matrix long title, series title, unit of measure and scalar factor). Always 5.
N2	I	4	Number of source cards (i.e. source and CANSIM series identifier). Always 2.
N3	I	4	Number of Note and Footnote cards. Up to a maximum of 21.
KSIGDM	I	4	Number of significant digits in DATABANK 10 if double precision 7 if single precision
KRTDEC	I	4	Number of digits to the right of the decimal point
ISPDP1	I	4	1 if double precision 0 if single precision
KTITLE	A	ST1x8	Matrix title of 300 characters followed by 20 blanks. Series title of 50 characters followed by 2 blanks. Unit of measure of 10 characters followed by 2 blanks. Scalar factor of 16 characters. Two lines of source information. Required number of lines of note and footnote information.
BUFI	I	4	Number of 4 byte words in DATA array.
KTAG	A	8	Same information as in LABEL.
M2	I	4	Always 1.
NSEC2	A	8	Same information as in NSEC1.
NOUT2	A	8	Same information as in NOUT1.
NOPEN2	A	8	Number of Characters in data record. 10 x BUFI.
DATA	S or D	BUFIx4	Data array. Contains BUFI single precision values or BUFI/2 double precision values.



PUBLICATION TAPE FORMAT

Matrix Record

Field	Length	Description
1-6	6N	Date: Date of Publication retrieval
7-12	6N	Matrix number
13-32	20	Series: Blank
33	1	Record type: 'M' (Matrix record)
34-35	2N	Record number: '99'
36-39	4	Agency responsible
40-43	4	Section responsible
44	1N	Crossfoot check
45-344	300	Long title
345-384	40	Short title
385-434	50	Source
435-934	500	Note
935-1054	120X	Footnote 1
1055-1174	120X	Footnote 2
1175-1294	120X	Footnote 3
1295-1414	120X	Footnote 4
1415-1534	120X	Footnote 5
1535-1654	120X	Footnote 6
1655-1774	120X	Footnote 7
1775-1894	120X	Footnote 8
1895-2014	120X	Footnote 9
2015-2025	11	(Not used)

Up to nine footnotes are allowed on each matrix, each footnote can be up to 120 characters.

Series Record

Field	Length	Description
1-6	6N	Date: Date of publication retrieval.
7-12	6N	Matrix number
13-32	20	Series number
33	1	Record type: 'S' (Series record)
34-35	2N	Record number: Last record is 99
36-39	4	Agency
40-43	4	Section
44-93	50	Title
94-103	10	Unit of measure
104-105	2N	Data mask type
106-108	3N	Variance allowed
109-110	2N	Scalar factor
111-112	2SN	Floating point characteristic
113-114	2N	Report frequency
115-117	3N	Expected time of update.
118-2021		Data points
	6N	Date of reference
	6N	Entry date
28	1N	Security code
	4N	Footnotes
	1N	Entry type
	10SN	Data point ¹
2022-2025		(Not used)

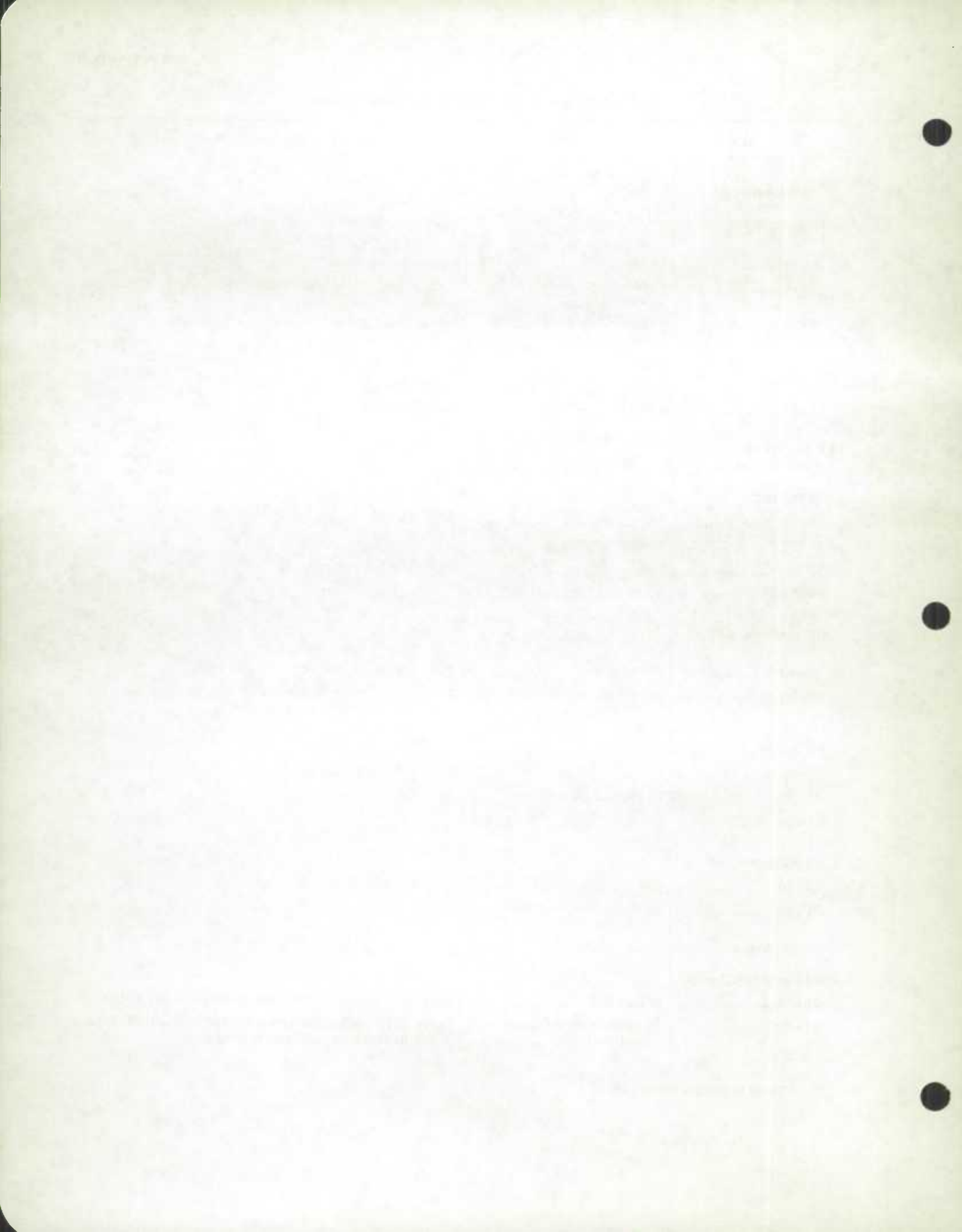
¹ There will be 68 data points on each series record.

Section	Acres	Original Grantee	Subsequent Grantee	Remarks
1	10
2	10
3	10
4	10
5	10
6	10
7	10
8	10
9	10
10	10
11	10
12	10
13	10
14	10
15	10
16	10
17	10
18	10
19	10
20	10
21	10
22	10
23	10
24	10
25	10
26	10
27	10
28	10
29	10
30	10
31	10
32	10
33	10
34	10
35	10
36	10
37	10
38	10
39	10
40	10
41	10
42	10
43	10
44	10
45	10
46	10
47	10
48	10
49	10
50	10

Card Format: ADD MATRIX, Operation Code AM

Column number	Contents	Explanation
Auto duplicate		
All cards ¹ columns 1-27:		
1 - 4	TSDB	System identification.
5 - 8	4 characters maximum, left justified.	Agency responsible for accuracy and security of data.
9-12	4 characters maximum, left justified.	Section of Agency responsible
13-19	Blank	
20-21	AM	Operation code.
22-27	6 digits	Matrix number.
Fields varying from card to card		
Card number:		
28-30	001	Card number.
31-51	Blank	
52	1 or 2	Crossfoot 1 = yes 2 = no.
53-80	Blank	
Card numbers 2-7 inclusive:		
28-30	002 to 007	Title card numbers.
31-80	50 characters maximum, left justified.	Title cards are continuous through 6 cards for a total of 300 characters.
Card number 8:		
28-30	008	Short title card number.
31-70	40 characters maximum	Short title.
71-80	Blank	
Card number 9:		
28-30	009	Source card number.
31-80	50 characters maximum	Source title.
Notes		
Card numbers 011 - 020:		
28-30	011 to 020	Note card numbers. One note is allowed per Matrix.
31-80	50 characters maximum, left justified.	Enter title continuously up to 500 characters. Do not use hyphens to continue to next card.

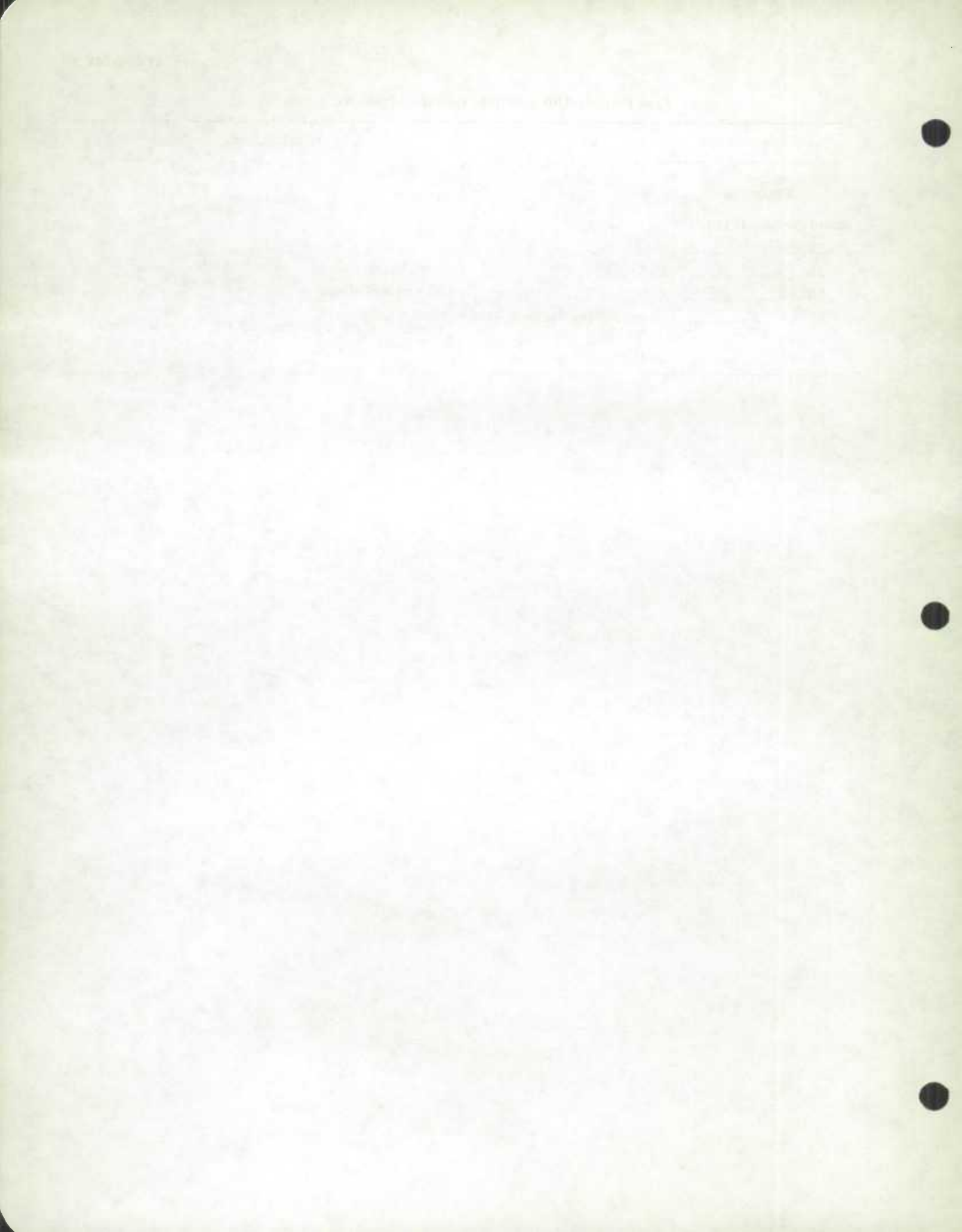
¹ There is no card number 10.



Card Format: ADD MATRIX, Operation Code AM – Concluded

Column number	Contents	Explanation
Footnotes		
Card numbers 111-193:		
28	1	1 = footnote
29	1-9	Footnote number
30	1-3	Footnote card number
31-80	50 characters maximum, left justified.	} Footnote text – a maximum of 120 characters through 3 cards.
31-50	20 characters maximum.....	

} Treated as
3-digit
card number



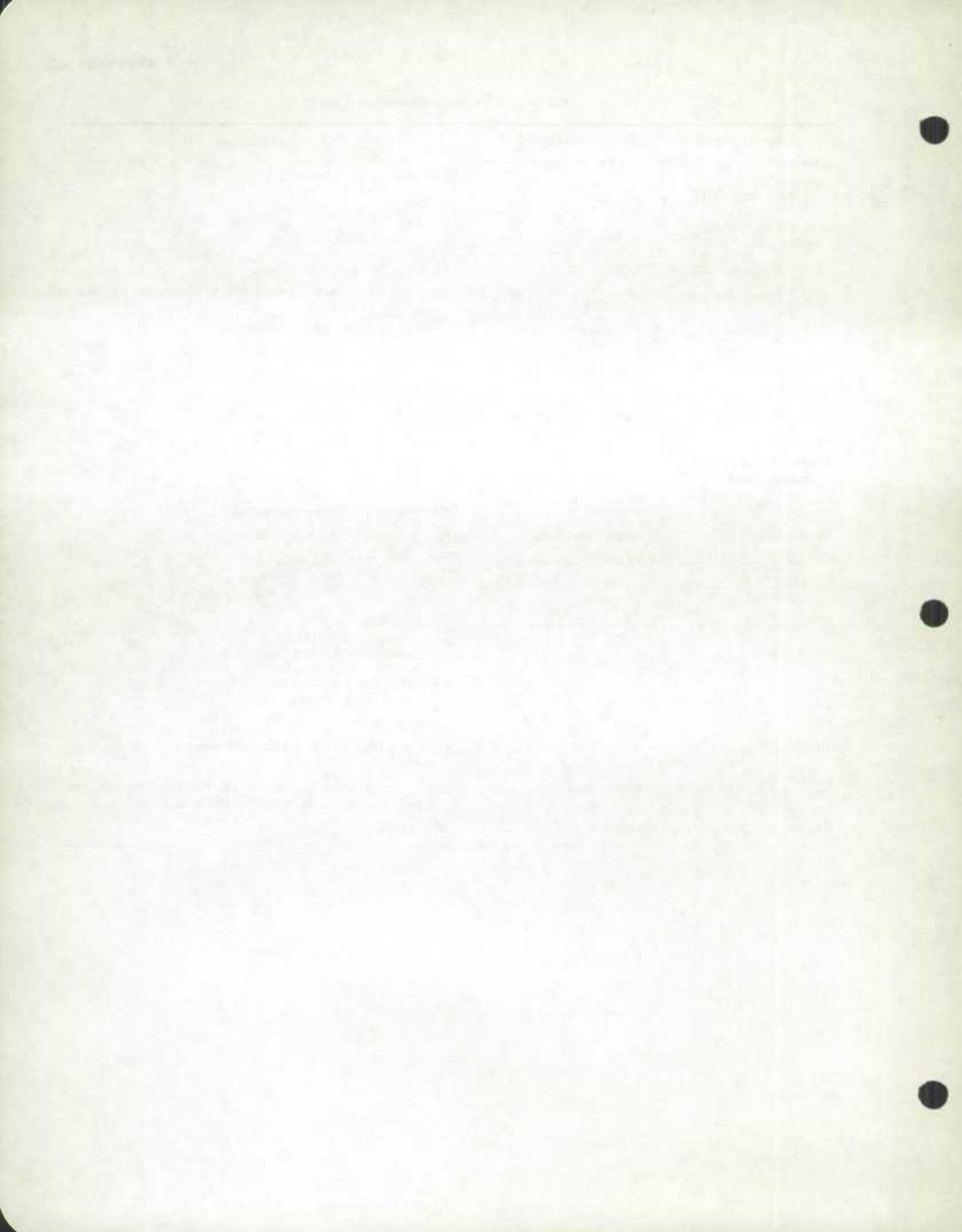
Card Format: ADD SERIES Operation Code (AS) Header

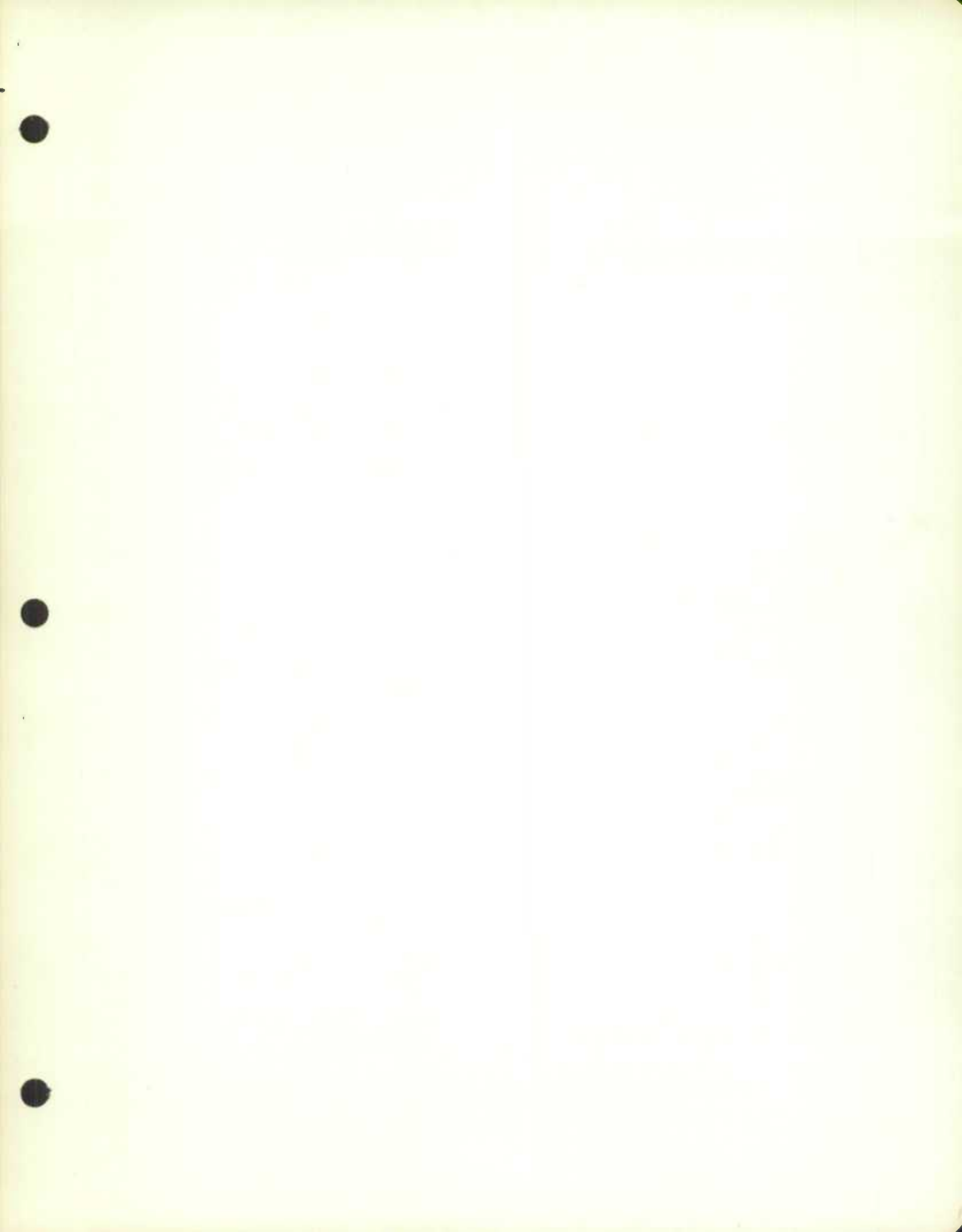
Column number	Contents	Explanation
Auto duplicate		
All cards columns 1-27:		
1 - 4	TSDB	System identification.
5 - 8	4 characters maximum, left justified.	Agency responsible for accuracy and security of data.
9-12	4 characters maximum, left justified.	Section of agency responsible.
13-19	Blank	
20-21	AS	Add series operation code.
22-27	6 digits, right justified	Matrix number, punch leading zeros.
Fields varying from card to card		
Card number 001:		
28-30	001	Card number.
31-50	20 digits maximum, left justified.	Series number.
51-52	00 to 12 or blank	Scalar Factor. Blanks are read as zeros.
53-54	- 9 to 12	Floating point characteristic.
55-56	00 to 99	Data mask type code.
57-59	001 to 998 or 999	Variance allowed, expressed as a per cent, as determined by the data source, or 999 = no edit requested.
60-66	Blank	
67-68	2 digit code	Report frequency.
69-71	3 digits	Expected time of update.
72-80	Blank	
Card number 002:		
28-30	002	Card number.
31-50	20 digits maximum, left justified.	Series number.
51-60	10 characters, left justified	Unit of measure, dollars, bushels, tons, etc.
61-80	20 characters, left justified	TITLE - first part.
Card number 003:		
28-30	003	Card number.
31-50	20 digits maximum, left justified.	Series number.
51-80	30 characters, left justified	TITLE - Second part.

DATE	DESCRIPTION	AMOUNT	BALANCE
10-11-54
10-12-54
10-13-54
10-14-54
10-15-54
10-16-54
10-17-54
10-18-54
10-19-54
10-20-54
10-21-54
10-22-54
10-23-54
10-24-54
10-25-54
10-26-54
10-27-54
10-28-54
10-29-54
10-30-54
10-31-54

Card Format: Enter Data, Operation Code (ED)

Column number	Contents	Explanation
Auto duplicate		
All cards columns 1-27:		
1 - 4	TSDB	System identification.
5 - 8	4 characters maximum, left justified.	Agency responsible for accuracy and security of data.
9 - 12	4 characters maximum, left justified.	Section of Agency responsible.
13 - 19	Blank	
20 - 21	ED	Operation code.
22 - 27	6 digits, right justified	Matrix No., punch leading zeros.
Fields varying from card to card		
28 - 30	001 to 999	Card numbers, to be sequential.
31 - 50	20 digits maximum	Series number, left justified.
51 - 56	6 digits	Reference date (yr., mo., day).
57 - 66	10 digits maximum, right justified.	Data.
67	1, 2, 3, 4, or 5	Type of data entry. 1 - Projection into future. 2 - Estimate of current figure. 3 - Current figure (update). 4 - Revision of current figure. 5 - Initial entry of data.
68	1 digit	Security level.
69 - 70	4 digits, maximum	Footnote indicators. A data point may have upto 4 footnotes.
73	Blank or 9	Blank if variance allowed will be checked by computer. For variance override, enter 9.
74 - 80	Blank	(Not used)





STATISTICS CANADA LIBRARY
BIBLIOTHEQUE STATISTIQUE CANADA



1010446877