A LIBRARY-BASED REFERENCE SERVICE FOR MACHINE-READABLE CENSUS DATA:

THE CANADIAN EXPERIENCE

Slavko Manojlovich

University of Windsor

Windsor, Ontario

The Census of Canada is a prime source for the demographic, social, economic and housing data used in social science research. Only a small proportion of the available census data is published in hard-copy form. If we add to this the growing use of computers for data analysis we can see that the machine-readable (M-R) version of the census is becoming increasingly more important to researchers. The library is the major depository of information on the university campus with the librarian acting as an available intermediary between the patron's informational needs and the library's resources. This paper will attempt to examine the place of the M-R census within the academic library. This will be followed by a discussion of the potential problems and pitfalls which librarians and researchers alike may encounter in making use of the M-R census.

THE M-R CENSUS AND THE ACADEMIC LIBRARY

The Census of Canada first became available in M-R form in 1961 with the production of the User Summary Tapes (U.S.T.). In 1971 the Public Use Sample Tapes (P.U.S.T.) were produced in response to the researcher's need for micro-data. For those libraries which had access to the M-R census tapes the referencing of the

patron's request for census information required a consideration, on the part of the librarian, of both the hard-copy and the M-R census resources.

The importance of the latter is evident in the following situations:

1. The requested data is not abailable in a printed publication but may be retrieved from a census tape. Since the census bulletins provide summary data for a limited number of standard geographic regions, they are unable to satisfy data requestes for small municipalities, subdivisions of these municipalities, or user-created geographic areas. City planning districts are an example of the latter. The major feature of the U.S.T. (enumeration area series) is that the record is based upon the enumeration area (E.A.). This is the smallest geographic unit of analysis allowed by the principle of confidentiality which governs the dissemination of census data. Furthermore, since E. A. boundaries never cross those of larger recognized statisti-

complete a description as indicated by the Statistics Canada information.

c) Furthermore, these respondents were asked, "Who processes the user's request for information from tape?". The majority response was the computer centre (33%), followed by the data

archive (20%) through the co-operative effort of the library and the computer centre (13). The remaining 33% responded with a variation of the user himself, with or without the assistance of faculty members in certain departments. (i.e. geography or sociology).

TABLE 1

Do you have any machine-readable census tapes on your campus?

SIZE OF HOLDINGS

		MAJOR	OTHER	TOTAL
Response	YES	8	7	56% (15)
	NO	7	5	44% (12)
				100% (27)

These observations clearly indicate that the overwhelming majority of academic libraries are not actively involved in providing user access to the M-r census resources which are located in the computer facility's tape library. Either the library is not aware of the existence of the M-R census tapes or, if it is, the processing of the data from tape is performed by a campus facility other than the library. According to Statistics Canada information, the M-R census tapes were rarely ordered by the library. Requests were made by either the computer centre or by individual faculty members. This

could possibly account for the library's position with respect to the M-R census. On the other hand, it is more likely that this is merely a reflection of the library's traditional indifference towards M-R information including data files and computer programmes. The addition of a chapter on M-R data files in the latest edition of the Anglo-American Cataloguing Rules (AACRII) should foster a willingness on the part of the library to accept the university's M-R resources as a legitimate part of the library's holdings.

transfer of entire data files to a patron's disc area by the data archive are both examples of secondary data reference. In either case, the patron is provided with the means for obtaining the desired information but not with the information itself. An attempt to expand their services to include primary data reference would not be reasonable in light of their need to acquire adequate support documentation.

The library has the personnel and the resources to provide a primary data reference service for the M-R census. The success of this venture depends on the following conditions:

1. The university community must recognize thelibrary as the resource centre for all census information, M-R and hard-copy. Thus, the library should be prepared to promote its expanded service by conducting several information and training seminars for faculty, staff and students on the availability and use of the M-R census.

2. The library will need to take an inventory of the M-R census tapes currently available on campus and should co-ordinate the future acquisition of tapes. Statistics Canada information indicated that 25% of the institutions with major holdings had acquired multiple copies of the tapes, an unnecessary added expense if you consider that the tapes are stored in the computer centre's tape library. At these

institutions tape requests were made by as many as 8 departments.

3. The majority of the libraries have a hard-copy computer terminal which they can use to acces the university's computer. However, the processing of user requests will require some staff training in elementary programming. An introductory course in a programming language such as PL/1 will more than satisfy the computer expertise necessary for massaging data from tape. A mini-session on the use of a statistical package such as SAS would also be helpful.

A library-based reference service for the M-R census has distinct advantages for the census user. First, the patron's request is directed to an integrated hard-copy and M-R census collection. Second, the request is processed by a reference librarian who is qualified in the art of query negotiation. The remainder of this paper will focus upon the problems which may arise in the referencing and use of the M-R census.

THE M-R CENSUS: PROBLEMS AND SOLUTIONS

Thelogical analysis of a typical census query into its component parts can provide us with a framework for comprehending the underlying problems associated with referencing census summary data. Statistics Canada has developed the PQRST classification as the basis of an indexing and retrieval system

TABLE 3

Total Population by Age Groups by Sex for Mersea Township, Ontario.

SEX	AGE GROUPS									
	55-59	60-64	65-69	70-74	75-79	80-84	85-89	90-94	95-99	100+
Male	175	165	170	135	90	50	20	0	0	0
Female	160	190	130	110	75	50	10	5	5	0
Total *1	335	340	305	230	140	70	5	5	5	0
Total *2	335	355	300	245	165	100	30	0	5	0

Reliability decreases →

* Total 1 is from the tape.

Total 2 represents my addition of male and female.

major realignment of boundaries. Witness the recent emergence of the "regional municipality" which replaced certain counties in the province of Ontario.

c) Code - A geographic area is identified in the M-R census via an assigned code. The introduction of the "Standard Geographical Classification" in 1975 produced a complete revision of these codes in the 1976 census.[16]

The identification of these potential changes is hampered by the fact that a geographic level of analysis may lie within the jurisdictional control of one of a number of government bodies. For example, E.A. and census tract boundaries are determined by Statistics Canada, whereas municipal boundaries are within provincial control. Since there is no single document which describes these geographic changes in the required level of detail one is forced to consult the appropriate publications of the respective government departments and agencies.

The · "Official Lists" of each census identify geographic areas in terms of E.

A.'s.[17]
Theoretically, this would enable a user to create a comparable area over census years. Unfortunately, correspondence lists for enumeration areas reveal that the nature of the boundary modifications do not permit the identification of equivalent units in all cases.[18] Statistics Canada does provide special tabulations for user-specified areas from the Census Master Data File should this become necessary.

d) Time: A major census is conducted every 10 years witha mini-census occurring at the 5-year interval. The next major census will be in 1981. It is sufficient to note that the mini-census contains only the core demographic questions and some additional housing or economic questions.

CONCLUSION

The first section of this paper described the current situation surrounding the accessibility of the M-R census at Canadian universities and made a case for the creation of a centralized census data reference service within the library. The discussion of the problems related to census data use underscored the need for a library or information professional to assist the researcher in the acqui-

ROUND II DELPHI QUESTIONNAIRE INSTRUCTIONS

PLEASE READ CAREFULLY

In the last issue of the <u>Newsletter</u> you were asked to respond to Round I of a Delphi study designed to aid our thinking about future developments relating to the distribution and archiving of machine readable data. That questionnaire was completely open-ended and was structured to elicit the ideas and concerns (in a number of categories) of the members of IASSIST. The responses from Round I were used in the development of the questionnaire presented in this issue -- a closed questionnaire consisting of forty-five "event" statements.

For this Round of the Delphi we are asking that you respond to each event statement with three different ratings or "Questions" (designated Question 1, Question 2, and Question 3). Questions 1 and 3 are seven point rating scales extending from "Low" to "High". Question 2 requests that you make an estimate of the Year in which the event will take place or begin. With these three Questions we are attempting to do the following:

1. Assess the importance of each event to the members of IASSIST (Question 1).

2. Forecast the approximate date by which the event will occur or begin to change (Question 2).

3. Estimate the likelihood that the event will actually occur (Question 3).

When the data are evaluated, median ratings on each of the three questions will be used as an estimate of the collective viewpoint of the IASSIST membership.

When using the seven point ratings scales for Questions 1 and 3 the scales should be interpreted as follows:

1. Very low importance or likelihood.
2. Moderately low importance or likelihood.
3. Low importance or likelihood.
4. Neutral importance or likelihood (a 50/50 chance).
5. High importance or likelihood.
6. Moderately high importance or likelihood.
7. Very high importance or likelihood.

To use the scales, simply CIRCLE the number which best approximates your viewpoint concerning the event statement.

Our primary concern is in forecasting directions in the nineteen eighties so your estimates of dates should generally be in the time period 1980 - 1990. It may be, however, that for a particular event you believe that the date will be later than 1990 -- if so, indicate the date. If you think the event will never happen, enter "9999" as the

===
```
        QUESTION 1            QUESTION 2           QUESTION 3
HOW IMPORTANT IS       ABOUT WHAT YEAR WILL  WHAT IS THE LIKELI-
THIS EVENT TO THE      THIS EVENT TAKE       HOOD THIS EVENT WILL
DELIVERY OF SERVICE?   PLACE OR BEGIN?       TAKE PLACE?

KEY   1  LOI=LOW IMPORTANCE         HII=HIGH IMPORTANCE
KEY   3  LOL=LOW LIKELIHOOD         HIL=HIGH LIKELIHOOD
```
===

1. A significant increase in the use of proprietary restrictions
    on the dissemination of data.

```
LOI                 HII                        LOL               HIL
                              YOUR ESTIMATE
  1   2   3   4   5   6   7      (       )      1   2   3   4   5   6   7
```

2. Expanded use of restrictions (copyrights, restrictive
    contracts) on the use of software used for archival and
    analytical purposes.
```
LOI                 HII                        LOL               HIL
                              YOUR ESTIMATE
  1   2   3   4   5   6   7      (       )      1   2   3   4   5   6   7
```

3. An expansion of the use of "non-standard" formats for
    disseminating data.

```
LOI                 HII                        LOL               HIL
                              YOUR ESTIMATE
  1   2   3   4   5   6   7      (       )      1   2   3   4   5   6   7
```

4. A decline in the ability of archivists to promote the use of
    standardized systems for defining data elements (variables).

```
LOI                 HII                        LOL               HIL
                              YOUR ESTIMATE
  1   2   3   4   5   6   7      (       )      1   2   3   4   5   6   7
```

5. A decline in the ability of archivists to promote the use of
    standardized systems for classifying and retrieving machine
    readable data.
```
LOI                 HII                        LOL               HIL
                              YOUR ESTIMATE
  1   2   3   4   5   6   7      (       )      1   2   3   4   5   6   7
```

6. A precipitous increase in the amount of machine readable
    data.

```
LOI                 HII                        LOL               HIL
                              YOUR ESTIMATE
  1   2   3   4   5   6   7      (       )      1   2   3   4   5   6   7
```

===
```
      QUESTION 1             QUESTION 2           QUESTION 3
 HOW IMPORTANT IS      ABOUT WHAT YEAR WILL  WHAT IS THE LIKELI-
 THIS EVENT TO THE     THIS EVENT TAKE       HOOD THIS EVENT WILL
 DELIVERY OF SERVICE?  PLACE OR BEGIN?       TAKE PLACE?

 KEY  1  LOI=LOW IMPORTANCE         HII=HIGH IMPORTANCE
 KEY  3  LOL=LOW LIKELIHOOD         HIL=HIGH LIKELIHOOD
```
===

13. Expansion of storage space requirements for data files.

```
LOI                 HII                        LOL                   HIL
                              YOUR ESTIMATE
  1  2  3- 4  5  6  7         (        )        1  2  3  4  5  6  7
```

14. Standardization of criteria for evaluating the importance of
    data files for inclusion in an archive.

```
LOI                 HII                        LOL                   HIL
                              YOUR ESTIMATE
  1  2  3  4  5  6  7         (        )        1  2  3  4  5  6  7
```

15. The rising cost of traditional methods for disseminating
    information (paper, printing) alter methods and techniques of
    dissemination.
```
LOI                 HII                        LOL                   HIL
                              YOUR ESTIMATE
  1 ,2  3  4  5  6  7         (        )        1  2  3  4  5  6  7
```

16. Expansion of the variety of users requiring machine readable
    data files.

```
LOI                 HII                        LOL                   HIL
                              YOUR ESTIMATE
  1  2  3  4  5  6  7         (        )        1  2  3  4  5  6  7
```

17. Expansion of   commercial archival services   for machine
    readable data.

```
LOI                 HII                        LOL                   HIL
                              YOUR ESTIMATE
  1  2  3  4  5  6  7         (        )        1  2  3  4  5  6  7
```

18. Development of more varied funding techniques for the
    development of machine readable data archives.

```
LOI                 HII                        LOL                   HIL
                              YOUR ESTIMATE
  1  2  3  4  5  6  7         (        )        1  2  3  4  5  6  7
```

===

QUESTION 1   QUESTION 2   QUESTION 3
HOW IMPORTANT IS   ABOUT WHAT YEAR WILL   WHAT IS THE LIKELI-
THIS EVENT TO THE   THIS EVENT TAKE   HOOD THIS EVENT WILL
DELIVERY OF SERVICE?   PLACE OR BEGIN?   TAKE PLACE?

KEY   1 LOI=LOW IMPORTANCE   HII=HIGH IMPORTANCE
KEY   3 LOL=LOW LIKELIHOOD   HIL=HIGH LIKELIHOOD
===

25. Increased need for access to machine readable data files
    through technical systems which are transparent to the end
    user.
   LOI   HII   LOL   HIL
                                   YOUR ESTIMATE
    1   2   3   4   5   6   7   (   )   1   2   3   4   5   6   7

26. Expanded training of potential end users concerning the
    availability and use of machine readable data files.

   LOI   HII   LOL   HIL
                                   YOUR ESTIMATE
    1   2   3   4   5   6   7   (   )   1   2   3   4   5   6   7

27. Dissemination of data files through networks rather than by
    tapes or other magnetic media.

   LOI   HII   LOL   HIL
                                   YOUR ESTIMATE
    1   2   3   4   5   6   7   (   )   1   2   3   4   5   6   7

28. Establishment of centralized and hierarchical data networks
    to improve access and to reduce cost of access to machine
    readable data.
   LOI   HII   LOL   HIL
                                   YOUR ESTIMATE
    1   2   3   4   5   6   7   (   )   1   2   3   4   5   6   7

29. Expanded use of microfilm or microfiche techniques in the
    dissemination of documentation and other relevant
    information.
   LOI   HII   LOL   HIL
                                   YOUR ESTIMATE
    1   2   3   4   5   6   7   (   )   1   2   3   4   5   6   7

30. Use of cassette tapes or diskettes used in conjunction with
    microcomputers (or with larger mainframes) for the
    dissemination of information.
   LOI   HII   LOL   HIL
                                   YOUR ESTIMATE
    1   2   3   4   5   6   7   (   )   1   2   3   4   5   6   7

===
<pre>
      QUESTION 1   QUESTION 2   QUESTION 3
 HOW IMPORTANT IS   ABOUT WHAT YEAR WILL   WHAT IS THE LIKELI-
 THIS EVENT TO THE   THIS EVENT TAKE   HOOD THIS EVENT WILL
 DELIVERY OF SERVICE? PLACE OR BEGIN?   TAKE PLACE?

 KEY 1 LOI=LOW IMPORTANCE   HII=HIGH IMPORTANCE
 KEY 3 LOL=LOW LIKELIHOOD   HIL=HIGH LIKELIHOOD
</pre>
===

37. Greater concern for problems of human engineering in the
    production of equipment designed for data retrieval.

```
LOI                 HII                       LOL                 HIL
                              YOUR ESTIMATE
 1  2  3  4  5  6  7           (       )       1  2  3  4  5  6  7
```

38. Development of software and hardware for dealing more
    automatically with natural language materials.

```
LOI                 HII                       LOL                 HIL
                              YOUR ESTIMATE
 1  2  3  4  5  6  7           (       )       1  2  3  4  5  6  7
```

39. Expanded deployment of font-independent readers for
    converting printed materials into machine readable form.

```
LOI                 HII                       LOL                 HIL
                              YOUR ESTIMATE
 1  2  3  4  5  6  7           (       )       1  2  3  4  5  6  7
```

40. Expanded concern on the part of machine readable data
    archivists with commercial and governmental data needs.

```
LOI                 HII                       LOL                 HIL
                              YOUR ESTIMATE
 1  2  3  4  5  6  7           (       )       1  2  3  4  5  6  7
```

41. Expansion of the scope of IASSIST to meet the demands of
    users other than researchers.

```
LOI                 HII                       LOL                 HIL
                              YOUR ESTIMATE
 1  2  3  4  5  6  7           (       )       1  2  3  4  5  6  7
```

42. Substantial resources devoted to the development of less
    expensive methods of information delivery.

```
LOI                 HII                       LOL                 HIL
                              YOUR ESTIMATE
 1  2  3  4  5  6  7           (       )       1  2  3  4  5  6  7
```

51. In what region do you reside?

    [ ] 1. Western Europe
    [ ] 2. Eastern Europe
    [ ] 3. Canada
    [ ] 4. The United States
    [ ] 5. Other (Where)_____

52. Institutional Affiliation:

    [ ] 1. Academic Institution
    [ ] 2. Commercial Organization
    [ ] 3. Governmental Agency
    [ ] 4. Independent data archive
    [ ] 5. Other (What)_____

53. Regardless of the type of agency for which you work, are you
    employed within a department which has:

    [ ] 1. Primary objectives other than data archiving?
    [ ] 2. Does primarily data archiving and servicing?

54,55. How many employees devote substantial amounts of time to
    data archiving? [_____]

                              Comments

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

17. Statistics Canada. <u>Official</u>
    <u>List</u>, 1971.

18. Statistics Canada. <u>Corres-</u>
    <u>ponding 1971-1966 Enumeration</u>
    <u>Area Numbers</u>,