

# Address Register Research for the 1991 Census of Canada

*Don Royce<sup>1</sup>*

**Abstract:** The current Canadian census is based on the local enumerator who lists and visits each dwelling in his/her enumeration area to drop off the questionnaires just before census day. Other approaches to compiling an address list exist, however. This paper describes current research into the use of an address register for the next Canadian census, to take place in 1991. Possible applications of an address register are described,

earlier research is summarized and potential sources for an address register in the Canadian context are mentioned. The paper concludes with a number of concerns which must be addressed in addition to the more technical issues.

**Key words:** Census; address register; coverage; administrative data; record linkage.

## 1. Introduction

One of the key activities in conducting most modern censuses is the creation of a list of dwelling units. In the Canadian census, this task is accomplished by first dividing the country into approximately 42 000 individual enumeration areas (EAs), each containing 200–300 dwellings. Each EA is then visited by a census representative (CR) who records the address of each dwelling in a visitation record (VR).

This address list has a number of uses in the conduct of the census:

- (i) it helps to ensure complete and accurate coverage of the EA by recording a unique description of each dwelling;

- (ii) it serves as a sampling frame to enable the CR to leave a longer questionnaire at every fifth household;
- (iii) it serves as a record of the status of each dwelling during collection, e.g., date questionnaires dropped off, date questionnaires received back, follow-up required, name of person to contact, etc.;
- (iv) it serves as a source of control information (e.g., person and dwelling counts) for subsequent processing of questionnaires; and
- (v) it serves as the source of interim dwelling and population counts.

In the Canadian census, the listing operation is done simultaneously with the delivery of the questionnaire. This combination has proven to be very effective, as it allows person-to-person contact with respondents, the deter-

<sup>1</sup> Methodology Branch, Statistics Canada.

mination of the number of questionnaires to leave in large households and the determination of the language (English or French) in which the questionnaire is needed. In rural areas the CR can identify farm households where an additional agriculture questionnaire is to be dropped off. It also allows the CR to observe and record data on dwelling type and permits a relatively efficient sampling methodology. It ensures that the list is up to date, as it is created during a short period just before census day, and it is believed to be a strong motivator for the CR, because he/she is responsible for the complete enumeration of the EA.

Creation of the basic address list by the enumerator at census time is not the only possible approach, however. The U.S. Bureau of the Census, for example, compiles address lists in advance of census day and delivers the questionnaires by mail. The address lists are purchased from private companies or are created by Census Bureau employees several months before census day. The lists are updated by a series of postal checks conducted up to and including the time of delivery of the questionnaires.

Many Western European countries have conducted their recent censuses using population registers. Sweden, for example, used a mailout-mailback technique in 1980 in which names and addresses were taken from the population registers.

At Statistics Canada, there are a number of reasons why investigation of alternatives to the current approach are of interest for the 1991 Census. For one thing, several of the advantages of the drop-off technique are declining in importance. Contact rates at drop-off are at 50 % and declining, so that in practice there is no difference between drop-off and mailout for a substantial portion of the population. Households themselves are getting smaller, and increased concern for personal security also means that gaining

physical access to some types of dwellings is becoming more difficult.

A second reason relates to coverage. The level of undercoverage has remained at approximately 2 % for the last several censuses, but with significant differential undercoverage between geographical areas and population subgroups. Publicity campaigns and various coverage improvement programs are used to keep undercoverage within acceptable limits, but these programs are very expensive. For example, a postal check conducted in 1981 cost \$1.4 million and added a total of 29 000 persons to the count. For 1986, this check was so expensive that it had to be cancelled. Alternative methods of compiling the basic address list which have the potential to improve coverage are therefore of considerable interest.

Third, the fact that the listing of dwellings is in manual form only limits the degree of automation that can be introduced in the collection and processing phases of the census. Examples of new techniques that could be used with an automated address list are described in Section 2.

The alternative discussed in the remainder of this paper is the use of an automated dwelling address register (AR). Section 2 describes the desirable features of an AR and potential applications if such an address register were available. Section 3 describes previous Canadian research on the topic. Section 4 discusses the current research plan for the 1991 Census, and Section 5 concludes with a number of important concerns in the development of an address register approach to census-taking at Statistics Canada.

## **2. Features and Applications of an Address Register**

The AR referred to in this paper would be a machine-readable file of residential addresses for some or all portions of Canada. Each

address would correspond to a dwelling unit, whether occupied or not, conforming to the Census definition. Each address in the register would also be coded to a precise geographic location. In urban areas this would normally be the block-face, while in rural areas it would correspond to the enumeration area.

In addition to these basic requirements, there are other desirable features of such an AR:

- (i) The record for each address would contain supplementary information, for example the type of dwelling unit (single, semi-detached, etc.), and the telephone number.
- (ii) The AR would be maintained on an ongoing basis throughout the five-year census cycle. Additions, deletions and changes to the file would be made continually as update information became available.

It is important to note, however, that the AR would not contain the names or any personal information about the occupants of the dwelling. Even though surnames would undoubtedly be useful, for example in resolving problems with multiple dwellings, it is felt that privacy considerations would make the inclusion of such information in the AR unlikely at the present.

It should also be noted that Statistics Canada has no such address registers at the present, although it has created and maintained experimental ARs in the past. Should an address register as described above prove feasible to develop, however, there would be a number of uses for it.

### *2.1. Mailout of the questionnaire*

With a reliable AR, the census questionnaire could be mailed out rather than dropped off. Comparative cost studies done as part of previous research (Booth (1976b)) forecast savings of between \$40 000 and \$652 000 for the 1981 Census and between \$569 000 and \$2.0 million for the 1986 Census (in 1976 dollars). These were net savings after deducting the cost of creating and maintaining an AR.

Most of the cost savings would come from the elimination of the drop-off phase of the CR's work, including the expense of training. In mailout areas, since staff would only be required for follow-up, a CR could handle a larger assignment; this makes possible an overall reduction in the number of field staff needed. This in turn would permit the hiring of better-quality enumerators.

### *2.2. Coverage improvement*

Even if a mailout census were not adopted, an AR could be used to improve coverage. One option would be to use the AR as an independent check on the address list compiled by the CR. Addresses on the AR but not listed by the CR would be checked in the field and would be added if they were found to be valid. Alternatively, the CR could be given a visitation record (VR) with the addresses pre-printed. The CR would then simply have to update the VR rather than create it. This methodology has the potential for both increased accuracy and lower cost (since listing would take less time). Such benefits would be expected even if the AR were simply the list of addresses from the previous census. If the AR could be kept up to date between censuses, the benefits would be even greater.

### *2.3. Delineation of field assignments*

CR assignments could be delineated on the basis of counts of the number of dwellings from an AR. An accurate AR could reduce the need for geographic field checks prior to the census and reduce the need for splitting EAs at census time, as is necessary when more dwellings are found in an area than anticipated. Availability of accurate counts from an AR would improve the overall efficiency of field operations by providing better information for planning the logistics of the collection operation.

#### 2.4. *Use of alternative sampling methods*

With an AR, the sample of dwellings which are to receive the long-form questionnaire could be pre-identified. By having more control over the selection of the sample, the potential for drop-off bias would be reduced. More complex methods could be used as well, for example varying sampling fractions and deeper stratification. Increased efficiency in sampling might allow a reduction in the overall sampling fraction and therefore a reduction in respondent burden.

#### 2.5. *Telephone follow-up of complete non-response*

In the 1986 Census, the mail return rate before follow-up is expected to be between 80 and 90 percent. With six million households in mail-back areas, this still leaves between 600 000 and 1.2 million households requiring follow-up.

Considerable time and cost savings can be achieved by contacting these households by telephone rather than field follow-up. In 1986, the CR will attempt to follow-up non-response households by telephone whenever possible. However, the success of this method depends on the CR's ability to obtain the name of someone in the household who can then be found for in the local telephone directory. Since contact at drop-off is only made about 50 percent of the time, this technique is of limited use.

An AR which also contained the telephone number for each address could increase this percentage considerably. Using the AR, a directory of telephone numbers could be created for each EA. A printout, sorted by address and showing the telephone number, could be used by the CR to locate the telephone number for any address at which a questionnaire was dropped off but not then returned.

For this application, the completeness and accuracy of the AR are not extremely impor-

tant. If an address is not found, then the CR will simply revert to the existing procedure.

A test of this application is being conducted in the 1986 Census in the city of Toronto, making use of a file purchased from the telephone company. Addresses on the file have been geocoded and printouts as described above are being provided to the CRs. A control sample of CRs who will receive no print-outs has been chosen and comparisons will be made of the rate of success of telephone follow-up between the two groups.

#### 2.6. *Centralized edit and telephone follow-up*

At present, the editing of questionnaires and subsequent follow-up operations are done by the individual CR for his or her enumeration area. However, centralized edit and telephone follow-up has been suggested for 1991, as it would allow better control of these operations and would also eliminate the "local enumerator" problem, where some respondents express reluctance to have their questionnaire examined by someone they know. The problem occurs particularly in rural areas, where the required questionnaires contain details of the farming operation and the CR may be from a neighbouring farm. However, the centralized approach has not been used to date, largely due to the logistical difficulties in coordinating the centralized activities with the more decentralized activities of the local CR (e.g., field follow-up).

An AR would serve as the basis for a file to monitor and control a centralized edit and telephone follow-up operation. The control file could be used for:

- (i) checking in questionnaires as they are received from respondents;
- (ii) recording the status of each questionnaire through the edit and follow-up operation;
- (iii) automatically creating and controlling assignments for both telephone and field follow-up;

- (iv) administrative functions such as production of management reports, calculating and verifying pay accounts; and
- (v) a source of data for subsequent analyses of the operation.

In effect, the AR would become a computerized version of the current visitation record.

### *2.7. Control file for census processing activities*

A control file based on an AR could be used not only in collection activities but also in many of the subsequent processing steps. For example, the file could be used to control the coding and data capture of questionnaires. The data from the "VR" (now automated) could then be immediately linked to the questionnaire. Many of the steps in subsequent processing which require access to the VR could be automated or in some cases eliminated.

In the long run, an automated AR could facilitate the integration of collection and processing activities. One could, for example, envisage a routine where the questionnaires are mailed back and then immediately captured (perhaps after some minimal pre-grooming), edited by computer, and batched for follow-up. Since the questionnaires are already captured, follow-up could even be by computer-assisted telephone interviewing (CATI). Once follow-up is completed, questionnaires which could not be resolved by follow-up would be sent for imputation. At the same time, interim population counts could be produced.

### *2.8. Integration of census and household surveys*

In addition to the census applications, an up-to-date AR could serve as the sampling frame for a wide variety of household surveys, including the monthly Labour Force Survey (LFS), post-censal or intercensal surveys linked to the census questionnaire, and inde-

pendent ad hoc surveys. An increased ability to conduct surveys would permit a better integration of data from the census with that from household survey programs. With increased demand for data on specialized subgroups of the population, one might move in the direction of using the census as a screening vehicle with more in-depth questions contained in a follow-up survey. The usefulness of an AR for this purpose would be greatly enhanced if the telephone numbers were also available.

Even if the addresses themselves were not used for sampling, an AR could serve as the basis for up-to-date counts of households for the design of ad hoc surveys or the update of continuing surveys such as the LFS.

## **3. Previous Canadian Research on Address Registers**

Research into address registers is not a new topic at Statistics Canada. As part of the testing programme for the 1971 Census, experimental address registers were created for two medium-sized cities, Kitchener-Waterloo and London, as described in Fellegi and Krötke (1967).

The Kitchener-Waterloo address register was created in 1966 from 1961 Census records, electricity billing lists and municipal assessment rolls. The assessment rolls were the best single source, containing 90 percent of the addresses on the joint list, followed by the electrical billings (83 percent) and the 1961 Census (65 percent). A dependent check by local letter carriers of the joint list resulted in 2 percent additions, although a quality control of this postal check (by withholding a sample of addresses on the list) estimated that the letter carriers missed a third of the addresses, resulting in an estimated undercoverage of 3 percent for the address register.

The London address register was created by field listing approximately one year before the test census day, using methods very similar to

those used for the monthly Labour Force Survey. The listing was quality controlled in the hard-to-enumerate parts of the city by re-listing a sample of city blocks. Two postal checks, one four months before census day and a second one just before census day, were done. The percentage of dwellings added by these checks was even higher than in the Kitchener-Waterloo study. For example, the first check added 4 percent, with an estimated one fourth of addresses missed by the letter carriers.

In the 1970s, a series of research studies was carried out on the feasibility of creating and updating an address register for major urban areas in Canada. A summary of these studies is given in Booth (1976a).

In one study, address registers were created for the three largest cities in Canada plus two smaller cities. For the three large cities the

registers were compiled by field listing by LFS enumerators. For the two smaller cities the address registers were created from the 1969 test census visitation records. The registers were updated and maintained for 18 months using data from a number of sources, including building permits and the post office. The ARs were then matched to 1971 visitation records and all non-matches were field-checked. Comparisons were made between the AR and the VR in terms of both undercoverage and overcoverage.

In terms of undercoverage, the two lists were quite comparable (Table 1). In the three large cities, the AR was deficient in picking up sub-addresses in the core of the urban areas. For single addresses, however, the AR was slightly superior to the VR. For the two smaller cities, the AR was superior for both single addresses and sub-addresses.

Table 1. Percent Undercoverage for Address Register Compared to Visitation Record

	Single		Type of dwelling		Overall	
	AR	VR	Sub-address		AR	VR
			AR	VR		
Three large cities	0.37	0.67	4.49	0.78	1.36	0.69
Two small cities	0.37	0.57	0.91	2.11	0.55	1.09

In terms of overcoverage, the AR was superior to the VR in the three large cities. In the two smaller cities the VR was slightly superior to the AR. Table 2 presents these results.

Table 2. Percent Overcoverage Rates for Address Register Compared to Visitation Record

	AR	VR
Three large cities	1.24	3.33
Two small cities	1.86	1.30

Subsequent studies concentrated on the question of whether an address register could in fact be maintained from just one source, namely "Point of Delivery" sheets from the Canadian Post Office. These are sheets which contain every basic address at which mail is delivered on each letter carrier's route. They are maintained in the local post office and are updated by the postal supervisor.

One of these studies was conducted in the 1974 Census Test city of Trois-Rivières. In this city, the AR was better than the VR both in terms of undercoverage (1.07 percent versus 2.93 percent) and overcoverage (1.55 percent versus 3.32 percent).

In summary, previous studies have demonstrated that an address register can be created and updated with information from just one source (the post office), and that the quality of listing in many cases can be at least as good as field listing. The one area of weakness seemed to be in sub-addresses in the larger cities. Despite the apparent potential of the AR approach, however, the capital cost involved was seen as problematic and the research did not continue.

#### 4. Research Plan for the 1991 Census

Since the studies carried out in the mid-1970s, there have been a number of developments which may make the creation and maintenance of an address register by automated means a viable proposition within the next five years. These include:

- (a) the availability or potential availability of an increasing number of up-to-date administrative files, such as records from telephone and electric utility companies, social security files, and tax files;
- (b) the universal use of the postal code on these files, which, combined with recently developed postal code-census geography conversion files, facilitates the automated restructuring of these files into census geography;
- (c) improved record linkage methodologies which could be used to unduplicate multiple files; and
- (d) the declining costs of storing and working with large databases.

Based on these considerations, a research plan was developed in 1985 which will permit key decisions to be made regarding the use of an address register in the 1991 Census by early 1989. The key steps in this plan are:

1. Carry out a comprehensive study of the various sources of and approaches to the construction and maintenance of an AR.

This would include an investigation of a possible joint venture with the post office to automate their lists.

Completion Date: September 1986.

2. Become more familiar with the experiences of other countries who use an address register and/or mailout approach for their censuses.

Completion Date: October 1986.

3. Construct a pilot register using an administrative data/record linkage approach. This step would include research into algorithms for standardizing and matching addresses so that duplicates in several files can be eliminated.

Completion Date: March 1987.

4. Update the cost comparisons done in 1976 to better estimate the potential cost savings with an address register – mailout census methodology.

Completion Date: June 1987.

5. Evaluate the quality of the pilot register in a post-censal study. Evaluation would focus both on the coverage of the AR and on the quality of data items on the AR. This study could be combined with other field testing for the 1991 Census.

Completion Date: June 1987.

6. If the mailout methodology appears to be feasible and promises sufficient cost savings, conduct a second field test to measure respondent and interviewer reaction, to estimate response rates, and to detect any problems with the methodology.

Completion Date: June 1988.

7. Update cost comparisons based on field test of mailout methodology and make a decision for 1991.

Completion Date: March 1989.

Whether steps 5 through 7 are carried out will of course depend on the success of steps 1 through 4.

Work to date has concentrated on the first of these steps. In 1985, a number of potential address sources were briefly investigated. These included the previously mentioned "Point of Delivery" records from the post office, telephone company files, municipal assessment lists, electrical company files, tax files, and listings from the Labour Force Survey and the 1981 Census. Although each has its advantages, it is quite clear that no single list is ideal. For example, the post office records are probably the most complete but they are currently unavailable in machine-readable form. Problems exist with electrical companies wherein apartment buildings are often metered as one unit. Telephone company billing lists contain the telephone number, which is useful for collection, but they exclude households without telephones and with unlisted numbers. Because no single list meets even the basic requirements listed in Section 2, the approach that is likely to be followed in the future will rely on the concept of a multiple-source address register unless it proves feasible to automate the post office lists.

Some research has also been conducted into address standardization algorithms. If large-scale address matching is to be done, methods to put addresses from several files into a standard form will be crucial. Testing done in early 1986 with existing algorithms and telephone company files identified several areas where improvements are necessary. Research into new methods is currently underway.

## **5. Concerns with an Address Register Approach for 1991**

In addition to the basic issues of cost, timeliness and quality which must be dealt with, a number

of other concerns must be addressed before the applications described in Section 2 could be implemented.

First, the use of a mailout methodology would make the census more vulnerable to errors, interruptions or cost increases in the postal service. With the present method, for example, a postal strike would severely disrupt collection, but not totally destroy it, since the CR could resort to pickup of questionnaires if necessary.

Second, what would be the public perception of an address register? The public may not distinguish properly between this and a population register, with all of its "Big Brother" connotations.

Third, a fully successful AR would require more geographical standardization than now exists. At the present, the census identification (enumeration area and household number) for the same dwelling unit changes from one census to the next. A geographic system which maintains the stability of geographical information down to the block face level would be needed.

Fourth, the coverage of the AR would have to be determined. At present, it appears to be much more feasible in urban areas than in rural areas. However, it is in rural areas where the potential cost savings are likely to be higher due to elimination of travel costs for drop-off.

Fifth, there is the question of how dependable the various sources of information needed to maintain the AR would be. For example, the post office has recently begun reducing the extent of home mail delivery as a cost-cutting measure. As a result, some addresses may not be available or may be difficult to obtain.

Finally, there is the question of how to minimize the risk in making a wholesale change-over from our traditional approach to one based on an address register. We would, in fact, be converting from an "area frame" approach to a "list frame" approach. Previous research, while encouraging, was based on a

relatively small number of cities, and it is likely that the testing described in Section 4 would also be limited in scope.

One possible solution is to create and maintain a full-scale national AR for 1991 without using it as the actual basis for enumeration. The quality of this AR would be evaluated during the 1991 Census by using it as a coverage improvement method. Dwellings missed by the CR but on the AR would be checked in the field and added to the census if valid. Dwellings listed by the CR but not on the AR would also be verified. Because this approach would compare two independently created lists, the evaluation of the quality of the AR is likely to be quite reliable.

In a second approach, the current drop-off collection methodology would be maintained, but the CR would be provided with the AR for his/her EA as a starting point for listing dwellings. Updates made by the CR would be data-captured and the quality of the AR evaluated by comparing the original to the updated version. However, since the verification of the AR is a dependent one, there is a risk that its quality could be overestimated. On the other hand, we would get valuable experience with the use of an AR in

the field and there may be the potential for some cost savings.

A third approach would be to use the full AR-mailout methodology in one or more test areas during the 1991 Census. The evaluation of this approach would require, at a minimum, a post-censal survey to measure coverage and a thorough analysis of response rates.

Whatever the approach, it is clear that the maximum potential of an address register will be realized only with a changeover to mailout of the questionnaires. The research program described in this paper will attempt to explore that question in the next few years.

## **6. References**

- Booth, J.K. (1976a): A Summary Report of All Address Register Studies Completed to Date. Statistics Canada report E-414E.
- Booth, J.K. (1976b): NAR Comparisons – Present Worth Approach. Internal Statistics Canada memorandum.
- Fellegi, I.P. and Krötke, K.J. (1967): The Testing Programme for the 1971 Census in Canada. American Statistical Association, Proceedings of the Social Statistics Section, pp. 29–38.

Received October 1985  
Revised July 1986