Administrative Data Use in a Traditional Census

Dolson, David
Statistics Canada
15-O R.H. Coats Building
100 Tunney's Pasture Driveway
Ottawa, Ontario, Canada
K1A 0T6
david.dolson@statcan.gc.ca

1. Introduction

The main purpose of this paper is to discuss ways in which administrative data are used in a traditional census. As it concerns Canada, this topic was first addressed by Carter and McClean (1996). Much has changed since that time, often in ways that were just taking shape at that time and which they either outlined or foresaw. In some regards this paper is an update on what they wrote.

First, what is a traditional census? Using a classification given in UNECE (2006) census-taking approaches can be classified as:

1. Traditional census
2. Traditional census with yearly updates of characteristics
3. Rolling census
4. Census employing existing administrative registers

A traditional census is one where data are collected with reference to a specific point in time (Census Day) and are obtained from a complete enumeration of the population and housing units. Various collection modes may be used. One questionnaire may be used or a long questionnaire may be administered to a (large) sample of dwellings with the remainder receiving only the short questionnaire. Canada is one of many countries using this approach.

In a traditional census with yearly updates, basic characteristics are collected periodically via a traditional census while an ongoing survey collects data for more detailed characteristics which can be pooled over time to produce estimates. The United States implements this method via its decennial Census and its American Community Survey.

A rolling census uses a continuous survey designed so that it covers 100% of the territory of a country over some period of time. Data are pooled over time to produce estimates. France is the only country using this methodology. It conducts a survey annually in the form of a census in one fifth of their small municipalities and a sample survey in their large municipalities.

Censuses employing existing administrative registers do so by linking together data from existing registers or other administrative sources. In some cases where the needed administrative data are not available or are of inadequate quality a complete enumeration or sample survey(s) may be used to produce required data. An increasing number of countries, mainly in Europe, are using register based approaches.

Obviously the fourth census approach makes extensive use of administrative records. However, the other three methodologies do also use administrative records for a variety of purposes and this paper will provide a description of how this is done in the case of a traditional census using Canada as the example.

As background to the main purpose of this paper, the next section notes, from a general perspective, potential uses of administrative records in a survey or census. This is followed by a short section identifying aspects of governance – legislation, policies and standard practices – that must be in place so that administrative records can be used. Section 4 gives a short outline of Canada's Census methodology for 2011,

noting points where administrative records are used. These are then discussed at more length in section 5, presenting each of the main uses including data sources, data quality, benefits and risks.

2. Using Administrative Data

"Administrative records are data collected for the purpose of carrying out various non-statistical programs" (Statistics Canada, 2009) and administrative purpose is defined by Canada's *Privacy Act* as the use of personal information about an individual in a decision making process that directly affects that individual including all uses for confirming identity and for determining eligibility of individuals for government programs. These statements imply records in government departments and agencies but for the purposes of this paper, I include also records maintained for similar sorts of purposes by private sector businesses.

They can provide a cost effective alternative to direct data collection and so reduce respondent burden and potentially provide a data quality advantage. There are several potential uses in a survey or census (Lavallée, 2007): for survey frames, whether directly or as a supplement to an existing frame; construction of sampling designs; replacement (partial or total) for data collection; in editing and imputation; direct tabulation; as auxiliary information in estimation; for survey evaluation. Most of these come into play in the Canadian Census program.

In using administrative records several important issues and risks must be considered. These include: accessibility to the records by the statistical agency, privacy concerns of those to whom the records refer, potential conceptual differences between the admin source and the statistical agency, data quality requirements and standards of the administrative data source viz-a-viz that needed for statistical purposes (e.g. coverage, timeliness, accuracy), the availability of suitable record linkage keys and importantly the legal/policy framework affecting the use of the administrative data. A final very important risk is that changes in laws and regulations can have serious ramifications for statistical use should a valuable administrative data source become unavailable or undergo major changes.

A further challenge in Canada is that many potentially valuable administrative sources are not federally administered and instead are provincial or even municipal in scope. When there is no intergovernmental committee or working group to address shared concerns and in particular that of coherence, issues related to concepts, variables included, file formats etc. can severely limit their use.

3. Legal, Policy and Governance Framework for Administrative Data Use

The governance of administrative data use by a national statistical office is a multifaceted topic, addressed in detail by Giles (2011). I provide here a short summary of the most important features.

A first requirement is legal authority to access and use administrative records for statistical purposes. Canada's *Statistics Act* provides Statistics Canada (StatCan) authority to obtain records maintained in any government department, corporation, business or organization; there is a mandatory requirement to provide such access when requested. This appears to grant StatCan unconditional access but it's not that simple. Sometimes legislation governing the administrative process has restrictions on access or secondary uses that in appearance or reality are in conflict with the Statistics Act. To address these issues and as a matter of good practice, it is the Agency's policy to negotiate this access and, if requested by either party, to set out the terms and conditions of access in the form of a Memorandum of Understanding. Amongst other things such agreements note: other relevant legislation or statutory authorities (e.g. legislation specific to the particular administrative data source), details of what data are to be provided including when, how often and how as well as its intended use, confidentiality protection and record linkage policy.

Information privacy is an important concern when administrative data are used for statistical purposes. Canada's *Privacy Act* ensures that the privacy of those to whom the data pertain is protected. As well, the confidentiality provisions of the *Statistics Act* mitigate against privacy invasion.

To produce desired statistical outputs administrative microdata must often be linked to other

administrative data, survey data or census data. This is certainly the case as it concerns the Canadian Census program. Such record linkages increase any invasion of privacy and are governed via Statistics Canada's Policy on Record Linkage which, amongst other things, ensures that the public value of each record linkage outweighs any intrusion on privacy that it represents. Any new record linkage activity requires careful documentation and senior management approval.

Canada has no universal personal identification number system. The Social Insurance Number (SIN) issued by Human Resources and Skills Development Canada, used in the administration of a variety of federal government programs, comes closest but its coverage is not universal and its usage at Statistics Canada is strictly limited mainly for reasons of privacy. Its most well known use is by the Canada Revenue Agency (CRA) in its administration of the *Income Tax Act* and related legislation. Many other identification numbers (e.g health insurance, drivers license) are used for other administrative purposes but there are no links between them or with the SIN. An important consequence for StatCan is that the very large majority of linkage work involving person records is done using probabilistic methods.

For major administrative sources with ongoing value to multiple Statistics Canada programs, specific organizational units are responsible on behalf of all users to negotiate and coordinate access with the holder of the administrative records. This approach has fostered more coherent, efficient and effective partnerships that benefit both parties and the Canadian public. The foremost example is the partnership between Statistics Canada and the Canada Revenue Agency, implemented at StatCan by the Tax Data Division (TDD). The CRA provides a wide variety of tax records to TDD which are used extensively in social statistics and especially in economic statistics programs throughout StatCan. Through this partnership it has been possible to implement at CRA data editing procedures and other steps of preparation for statistical use of the records. Once at StatCan, further steps of preparation are completed and staff of the TDD act as the subject matter expects for tax data. The TDD then is responsible for stewardship of the data both as it concerns the *Statistics Act* and StatCan policies but also on behalf of CRA to ensure that provisions of the *Income Tax Act* and other tax legislation are respected. They must also track and control access within Statistics Canada. Yung et al. (2011) provide an extensive discussion of tax data use at the agency.

Other major examples germane to the Census program where StatCan undertakes a centralized coordination approach are vital statistics records from provincial/territorial vital statistics registries and immigration records from Citizenship and Immigration Canada. Less broadly used administrative data sources are managed and coordinated on a case by case basis.

4. The Canadian 2011 Census and National Household Survey

Although the Canadian Census remains a traditional Census its strategy is considerably changed, both in 2006 and in 2011, from that used in the previous several Censuses.

Starting in 2006, the majority of questionnaires have been delivered to dwellings via Canada Post Corporation, using Statistics Canada's Address Register (AR). The AR is maintained through feedback from the Census itself, an evolving variety of administrative sources and field listing/verification procedures.

In 2011 about 80% of dwellings – those in areas where the AR coverage and quality were judged to be sufficiently high and where Canada Post Corporation uses the address for mail delivery - were initially contacted via mail and another 18% via dropoff of a questionnaire. Mail delivery and dropoff took place mainly in the week prior to Census day, May 10, 2011. The remaining 2%, primarily in remote locations, were interviewed in person. Amongst those receiving their "questionnaire" via mail three quarters received an Internet promotion letter, providing a unique access code linked to the address and encouraging response via Internet; instructions were also provided on how to obtain a paper questionnaire if so desired. This area was identified via detailed demographic analysis of Internet responses to the 2006 Census as well as paradata from the 2006 Census (local response rates, response mode rates) both in general and of data from an embedded test of the Internet promotion letter strategy to identify those areas with the highest probabilities of response via Internet.   The remainder of the mailout group as well as those in the dropoff area received a

questionnaire package providing both a paper questionnaire and an Internet access code.

In a Wave Methodology described in detail by Laroche (2011), a number of additional waves of contact were delivered at carefully timed intervals with the goal of maximizing self-response preferably via Internet or alternatively by mail. For the first time anywhere Internet response to the Census exceeded 50%. As usual, follow-up of non-response culminated with field follow-up by enumerators.

Up to 2006 80% of dwellings received a short questionnaire including only basic demographic characteristics and the other 20% got a long questionnaire collecting both the basic characteristics and information on a series of topics such as language, employment, occupation, income, disability status, etc. In 2006 for the first time in the Census, long form respondents were offered the option of having Statistics Canada retrieve their income information from tax records instead of answering the income questions.

In a major departure from the past, in 2011 collection of the more detailed long form data was made voluntary and the National Household Survey (NHS) was created. The Census remained mandatory and, using the data collection methodology described above, every dwelling received a short questionnaire including only basic demographics and some language questions. The NHS with a sample size of about 30% of dwellings remained part of the Census program. Its questionnaire included the Census questions as well as content previously on the Census long form. Collective dwellings continue to be covered in the Census but are out of scope for the NHS.

NHS data collection was generally similar and very integrated with that of the Census. Dwellings in the NHS sample who responded to Census via Internet were immediately offered to continue and respond to the NHS as well. Those who responded to Census via mail prior to the end of May were mailed a NHS questionnaire (Internet access code also provided). Finally, those who did not respond to Census by that time were not delivered a NHS questionnaire but rather were interviewed for NHS as well at the time of their non-response follow-up interview for Census.

During Census/NHS processing a major item will be weighting and estimation for the NHS and calibration of those estimates to Census counts. Research for this is ongoing but it is very likely that administrative data in the form of tax records will play an important role.

Other typical activities of processing such as data capture, editing, imputation for item and unit non-response will be completed over the coming several months. As processing is completed for each module of data and prior to each release date, subject matter experts will validate the data for internal consistency and confront counts and estimates with figures from other surveys and administrative sources.

A final component of the Census program is the coverage studies (Dolson, 2010). These use a wide variety of administrative records as frames, for tracing and in estimation.

5. Administrative Record Use in a Traditional Census

It seems likely that virtually every traditional census would use administrative data at some point in its operations whether in planning, data collection, evaluation or somewhere in between. In a modern (largely) traditional census such as that conducted in Canada administrative data are used at several stages:

1. Although it can be debated whether they are indeed administrative data, paradata from the 2006 Census were used heavily in planning for 2011.
2. In maintaining an Address Register -- used to create a list frame for the Census
3. To reduce respondent burden in data collection and to improve data quality by using tax records for income information for respondents who indicate permission for such use
4. To replace direct data collection from residents of certain collective dwellings
5. In maintaining a Business Register, used to assist in coding of long form industry data
6. In the NHS (Census long form) by using tax records to provide auxiliary data to assist in estimation
7. In data quality evaluation of Census/NHS counts and estimates prior to their release
8. In frame creation, respondent tracing and estimation in the Census coverage studies.

Each of these is now discussed in turn.

5.1 The Previous Census

It is entirely normal to use the results from one iteration of a survey or census to help plan and develop the next. With the rapid expansion in the availability of paradata produced in the administration and management of data collection activities the scope of potential such applications has also expanded. Here, I briefly note two uses of what might be called administrative data from the 2006 Census – paradata from the Census' data collection management system – in planning the 2011 Census.

First, to determine the areas of the country to receive the Internet promotion letter rather than a questionnaire package required extensive linked analysis of 2006 Census paradata and data. In particular, a demographic profile was derived for those dwellings responding to the Census by Internet. Using this, it was then possible to identify those areas of the country with the greatest likelihood of response via Internet. These areas ended up accounting for about 60% of the residential addresses in Canada.

Second, development of the wave collection methodology critically depended upon detailed analysis of the timing of receipt of responses by each response mode. This facilitated extensive modeling and testing of alternate timings and modalities for each wave of respondent contact for 2011.

5.2 The Address Register

Statistics Canada's Address Register (AR) was originally developed for use in the 1991 Census as a coverage improvement tool and was created using records from four administrative sources: personal income tax records from the CRA, telephone company billing files, electricity company billing files and municipal property tax records. The AR was used to update enumerator created lists of addresses and vice versa. After dropping off the 1991 Census questionnaires, enumerators compared the list of addresses in their visitation records against a listing of addresses taken from the Address Register and any discrepancies between the lists were investigated. Valid dwellings identified by the Address Register were then enumerated in the Census.

This use of the AR was a success and was repeated in the 1996 and 2001 Censuses. During this period there was some evolution in the choice of administrative sources and considerable improvement in address matching methods (to remove duplicates between the differing sources) and geocoding techniques.

With steady improvement in its quality and geographic coverage the AR included addresses for about 90% and 95% of the dwellings in Canada prior to the 2006 and 2011 Censuses, respectively. However, not all of these could be used in a list frame for mailout in the Census. Only those areas with both civic style addressing to which Canada Post Corporation could deliver mail and where there was adequate quality of address coverage were retained for the mailout frame for the Census. This included 67% and 80% of residential addresses in 2006 and 2011, respectively.

Although AR maintenance via administrative sources does an excellent job, it alone is not sufficient to ensure sufficiently high quality. Thus to be assured of a complete and accurate list of addresses for the mailout area, each of these Censuses was preceded by a field verification operation – on a 100% basis for 2006 and on a targeted basis covering about 40% of mailout collection units for 2011.

Statistics Canada's household survey – particularly the Labour Force survey - program started using the AR in 2004 to assist in cluster listing activities. Moving towards the future this use is expected to expand considerably. Consequently the AR is now maintained continuously with an ongoing program of quarterly updates from both administrative sources and field listing; McClean and Charland (2011) give a detailed description. Currently the administrative sources used are:

- Telephone company billing files
- Commercial telephone directory files
- Tax files from the CRA (personal income tax records, now housing tax rebate records, Canada Child Tax Benefit records)
- Canada Mortgage and Housing Corporation Starts and Completions Survey records

- Canada Post Corporation Point-of-Call file
- Various municipal sources

These are received variously from monthly to occasionally and have differing levels of quality, completeness and timeliness, all of which can vary over time. So an ongoing program of quality assessment is essential not only for each source but also in terms of how well they work together.

The list of administrative sources will change over time. For example, the Point-of-Call file is a relatively new addition. This is a list of addresses to which Canada Post Corporation delivers mail. Not surprisingly, it has turned out to be a very timely source with excellent coverage. On the other hand with the rapid increase in cell phone only households telephone company billing files, although still useful, are not as good a source for AR updates as they once were. Municipal sources differ widely in their quality, ease of use and marginal utility given the other sources with broader geographic coverage.

Ideally the AR should include a unique address for each dwelling in Canada. But there are a number of challenges having the impact that it does not. Apartment units, particularly in small buildings, are a significant one. Address standards requiring unique identifiers do not exist. Different administrative sources variously do not provide apartment unit labels, use different labeling (e.g. apartments 1, 2 and 3 versus A, B and C) or impute them. Whether or not to include a basement apartment, say, may not be clear cut and can depend on interpretation by the respondent or enumerator. This inconsistency can cause both duplication and undercoverage on the AR due to the difficulty in address matching. Both administrative sources and field listing have difficulty detecting these changes let alone in a timely way. Overcoverage also results when streets are known by more than one name, with different variations appearing on different administrative sources. Given the lack of a source indicating these aliases, address matching has great difficulty detecting these duplicates.

5.3 Tax records to replace survey data

For respondents who indicated to do so, income tax data taken from records from CRA were used for direct replacement of data in the 2011 NHS that would otherwise have been collected from the respondent. After successful development and use of this methodology with the Survey of Labour and Income Dynamics (Michaud et al., 1995) it was first implemented for the Census in 2006 and again used in the NHS in 2011. In particular, respondents are first asked if they would like Statistics Canada to use their income tax records to extract income data for the Census/NHS. Those who reply "yes" then skip the income questions while those responding "no" proceed through a detailed series of income questions.

These questions impose a substantial respondent burden and historically have suffered from item response rates in the order of only 50%. (It should be noted though that for income items – and a number of others as well - an apparent item non-response may easily be an implicit reporting of a zero amount or not applicable). Further, responses given are frequently rounded or estimated. The Census-tax linkage approach to collecting income data for Census/NHS has proven to have major benefits in response burden reduction and in data quality improvement.

On the 2006 long form approximately 82% of respondents granted permission to use their income tax data. The record linkage from Census records to income tax records was done using probabilistic methods. Excellent quality linkage variables such as name, date of birth, sex, marital status, spouse name and address were available and so a tax record match was found for 88% of those who had granted permission. Evaluation studies indicated a false match rate of only 0.5%.

In the 2001 Census there was a great deal of rounding with more than 50% of reported income amounts ended in 000; in contrast this figure was only 9% with use of the tax data in 2006. A second important improvement was much increased reporting of small amounts that were either omitted or forgotten when respondent reported.

5.4 Collective Dwellings

A collective dwelling is a dwelling used for commercial, institutional or communal purposes such as a hotel, a hospital or work camp. Other examples include rooming houses, nursing homes, university residences, military bases and correctional facilities. In many of these, such as hotels and university residences, residents complete their own census forms (paper questionnaire only). In institutional collectives such as nursing homes or correctional facilities, where residents are unable to respond on their own behalf or there are safety concerns, enumeration information is taken from administrative records maintained by the collective. Each such collective must be contacted but residents are not asked to complete questionnaires.

In a first step of data collection a profile form is completed for each collective dwelling by interviewing a representative from the management of the collective. Information from this allows, amongst other things, confirmation or determination of the collective's classification and thus the determination of the appropriate data collection methodology. The actual enumeration is performed closer to census day by an enumerator visiting the collective and either leaving forms for residents to complete (to be picked up later by an enumerator), or completing forms for usual residents from administrative records.

Up to the 2006 Census most non-institutional collectives received the long questionnaire while institutions got only the short form. In 2011 all collectives are in the (short form only) Census and are out of scope for the NHS.

While these administrative data are usually of good quality this is not always the case. Although the coverage of these administrative records is generally excellent, often we are able to collect only name, sex, age and date of birth. Marital status and common law status frequently require imputation. Also, the expanded set of language questions now on the short form for 2011 seems likely to require high rates of imputation when data are collected from administrative records at these institutional collectives.

5.5 The Business Register

In the NHS, respondents aged 15 years and over are asked "For whom did this person work?" as well as the address and the kind of business activity; answers for all of these are used in determining the appropriate industry code. The major reference file, derived from Statistics Canada's Business Register (BR), indicates the industry code for every business establishment in Canada. Although there is ongoing profiling and classification work undertaken by Statistics Canada, the major source of information on the birth of businesses and their industry is obtained from administrative records obtained from the CRA. Each time a business registers a new payroll deductions account with CRA this can indicate the birth of a new business or a new establishment within an existing business. These records which indicate the name, location and planned activity are forwarded to StatCan monthly. As appropriate businesses are birthed to the BR and assigned an industry code. For the Census this provides the basis for industry coding and thus a variety of statistics by industry that will be coherent with those produced by other programs at the agency.

5.6 Tax records as auxiliary data

The estimation strategy for the long form variables collected via the NHS in 2011 will very likely use income tax records as auxiliary data.

Because it is voluntary, a substantial portion of the NHS sample required non-response follow-up (NRFU). Given Census/NHS resource and time constraints a consequence was that after an initial period of follow-up for all NHS non-respondents, a subsample of remaining non-respondents was selected for continued followed-up. The result is a rather unique two phase sample design. This is unique because the integrated approach to Census/NHS data collection ensures that basic demographic and language data collected via the Census are available on a near 100% basis for all dwellings in the NHS sample whether they responded to the NHS or not and whether in the NRFU sub-sample or not. Thus, in its first step, estimation for the NHS can be viewed as a partial imputation problem where the long form data are imputed for non-respondents in the NRFU sub-sample.

In a second step the weights of all units in the sub-sample are adjusted to account for the NHS non-respondents not included in the sub-sample. In a third step NHS weights will be calibrated via a regression estimation procedure such that NHS estimates for the Census variables will be equal to the Census counts for each of about 6,000 weighting areas.

The imputation will be done using a nearest neighbor methodology with donors selected from the pool of all NHS respondents. To enhance the quality of this imputation strategy and so that of NHS estimates it is feasible to use as matching variables not only the demographic and language variables but also income variables which can be obtained from tax records using the same linkage strategy as that described above in section 5.3. This is particularly valuable given how strongly income is associated with many other long form characteristics. Note that the question of obtaining respondent agreement for use of their tax data is moot since these data will be used as auxiliary variables will not be retained on the database. Although not all details of the estimation strategy are yet worked out, testing and simulations so far indicate a clear data quality improvement from use of these administrative data in this way.

5.7 Census data certification

During the period leading up to each Census or NHS publication date, counts and estimates are thoroughly reviewed by subject matter experts in terms of internal consistency, comparability with results from the previous Census and for their coherence with external indicators. This evaluation is usually done at higher levels of aggregation and depending on the particular topic the external indicators could include one or more of results from other Statistics Canada surveys and data from a variety of administrative sources. Often there are conceptual, coverage and other differences analysts must take into consideration as they conduct their evaluations. Some examples follow.

Estimates of the population of Aboriginal persons are compared to counts available from Aboriginal Affairs and Northern Development Canada. This can be a difficult comparison as aboriginal identity is a fluid concept which is self-reported. Further, there is potentially large undercoverage on administrative files depending on the aboriginal group.

Census estimates of intercensal immigrants to Canada are compared to counts from Citizenship and Immigration Canada for such characteristics as place of birth, year of immigration, location in Canada. There are no great difficulties with concepts but comparisons by geography must be done with care as locations provided on immigration records are intended addresses or locations and often differ from actual locations of settling.

Estimates of one and five year mobility can be compared to figures derived from tax data from CRA. There are some conceptual and coverage challenges. For example, children do not file tax returns and so their mobility is assumed to be the same as their parents. Even amongst adults, not all file tax returns.

5.8 The Coverage Studies

Census coverage studies in Canada use the Reverse Record Check (RRC) methodology to measure undercoverage while overcoverage is measured by matching the census database to itself (Dolson, 2010). Both make extensive use of administrative records.

The RRC is a multiple frame survey in which three important components are derived from administrative sources. The frame for intercensal births is taken mainly from vital statistics records. These are considered to be of very high quality with essentially 100% coverage but unfortunately their timeliness is not entirely adequate. For the 2011 RRC, the frame for 2010 and 2011 births will be taken from Canada Child Tax Benefit files originating from the CRA. These files are available much more quickly and our evaluations indicate their quality and coverage is also excellent. Next, the frames for intercensal immigrants to Canada and for persons with permits for temporary residence in Canada are taken from immigration records from Citizenship and Immigration Canada. Coverage in these files is excellent and their timeliness is good. However information from these files used to locate and interview sampled persons is often of poor

quality making tracing work difficult. Finally, frames for the three northern Territories are taken from Territorial health care files, obtained by special arrangement specifically for use only in the coverage studies. These files are maintained on a very current basis and do well at keeping up to date relative to the high rates of in and out migration experienced in the territories.

Persons selected into the RRC sample must be traced forward from their last known address. For many, this address is five years old – the location where they were enumerated in the previous census. Sampled persons are traced forward first using income tax records. Record linkage is probabilistic using name, spouse name and demographic variables for linkage; updated addresses are successfully obtained for a large percentage of the sample. Further tracing is done for a large number of sampled persons using drivers license records obtained (indirectly) from provincial ministries of transportation; these are an excellent source of current addresses. Numerous other more local administrative sources are used by data collection staff as required in tracing more difficult to contact persons in the RRC sample.

The Census Overcoverage Study matches the census database to itself but does so indirectly by first matching it to a file of administrative records constructed from several years worth of birth registrations and income tax records. A complete description of the methodology is given in Morel and Farr (2007). Several steps of probabilistic record linkage of persons to persons and of households to households are undertaken in this work using name, family composition and demographic variables for linkage. The availability of names on the census response database and the use of the administrative records were critical in facilitating this methodology, resulting in much improved precision in estimates of overcoverage. In addition to some computational advantages an important quality advantage was that matching first to the administrative records made it much easier, for example, to distinguish between the correct enumeration of two individuals with the same name and date of birth versus duplicate enumerations of one of them.

6.　Conclusion

As an example of a traditional census, the Canadian Census has surprisingly many features where administrative data play a role. There are only two situations where direct use of administrative data takes place: tax records in lieu of respondent reported income data in the NHS; and administrative records in certain types of collective dwellings for Census variables. The other uses are no less important but are indirect in nature. In every case use of the administrative records has resulted in one or more of reduced cost, reduced respondent burden, improved quality and better assessment of quality.

REFERENCES

Carter, R. and McClean, K. (1996). Using Administrative Data in the Canadian Census: Experiences and Plans. Presented at ECE/Eurostat Joint Work Session on Administrative Registers and Administrative Records in Social and Demographic Statistics, Luxembourg.

Dolson, D. (2010). Census Coverage Studies in Canada: A History with emphasis on the 2011 Census. Proceedings of the Section on Survey Research Methods, American Statistical Association.

Giles, P. (2011). A Canadian perspective on the use of administrative information for statistical purposes. Presented at the 58th ISI Congress, Dublin, Ireland.

Laroche, D. (2011). The Evolution of Edits in the Canadian Census of Population Online Questionnaires. United Nations Statistical Commission and Economic Commission for Europe. Work Session on Statistical Data Editing, Ljubljana, Slovenia.

Lavallée, P. (2007). Administrative Data Usage in the framework of Social Statistics:Current and Future

Picture. Internal report, Statistics Canada.

McClean, K. and Charland, K. (2011). Address Register: Future State after the 2011 Census of Population. Internal report. Statistics Canada.

Michaud, S., Dolson, D., Adams, D. and Renaud, M. (1995). Combining administrative and survey data to reduce respondent burden in longitudinal surveys. Proceedings of the Section on Survey Research Methods, American Statistical Association.

Morel, J. and Farr, H. (2007). Measuring Person Duplication: The 2006 Canadian Census Overcoverage Study. Proceedings of the Section on Survey Research Methods, American Statistical Association.

Statistics Canada (2009). Statistics Canada Quality Guidelines. Catalog 12-539-X.

UNECE. (2006). Conference of European Statisticians Recommendations for the 2010 Censuses of Population and Housing. ECE/CES/STAT/NONE/2006/4, United Nations Publication, ISSN 0069-8458.

Yung, W., Leblanc, D. and St.-Louis, G. (2011). The Use of Tax Data in Official Statistics – The Canadian Experience. Presented at the 58th ISI Congress, Dublin, Ireland.

ABSTRACT

Methods used in traditional censuses have advanced considerably in recent cycles and some now make considerable use of data from administrative sources, although primarily in an indirect way. Countries such as Canada and the U.S.A. now mail out the majority of their census questionnaires using high quality address registers whose maintenance is based heavily upon administrative data. The closely related census coverage studies and population estimation programs benefit from several administrative sources including vital, tax, immigration and other records. In Canada's 2006 Census, income tax data were use directly in the Census long form for those respondents who agreed to its use for this purpose. In the 2011 National Household Survey, which replaces the long form, this will again be done. The tax data will also be use in estimation procedures designed to minimize the impact of non-response bias. Focusing mainly on Canada as an example this paper will outline these uses and discuss related issues of access, data quality and its measurement as well as the potential for extended use of administrative records in the future evolution of traditional censuses.