

Integration of Administrative Data with Survey and Census Data

Michael Trant

Statistics Canada, Agriculture Division

Patricia Whitridge

Statistics Canada, Business Survey Methods Division

ABSTRACT: Statistical systems are a consequence of evolution, and the level of data integration that is achieved is often an indication of the degree of system development. Almost all national statistical agencies in the world integrate administrative data with their survey and census information to some degree in order to complement, supplement or replace survey information, or to assist with frame maintenance. This paper reconstructs the development and evolution of the Canadian agricultural statistical system as it relates to the expanding and increasingly important role of administrative data.

The degree to which administrative data are integrated depends on a number of factors, the most important being: (1) the degree of maturity of the country's statistical system, (2) the quality and the amount of information available from the government's administrative and regulatory programs, (3) well-trained and experienced staff, (4) funding, and (5) cooperation among government agencies. Most countries appear to have gone through an evolutionary process in establishing their agricultural statistical system and most systems are developed with an internal capacity for renewal and adjustment. This allows them to respond to changing conditions and needs, and to remain relevant. The actual route that is followed, however, is highly dependent on the amount of resources available for the program, the availability of experienced professionals to develop and maintain the system, and the statistical toolbox that they are able to use.

1. Introduction

Almost all the national statistics agencies in the world integrate their administrative data to some degree with their survey and census information. The focus of this paper is a description of the statistical initiatives used in Canada to search for banks of administrative data to replace or complement what has traditionally been collected using survey and census information. Much of the statistical information in any national statistical system comes from administrative sources. Statistics on births and deaths, immigration and emigration are almost always compiled from administrative sources. Similarly customs documents on imports and exports form the basis of information on trade in agriculture and food products. This paper reconstructs the development and evolution of the Canadian agricultural statistical system as it relates to the integration of information and the expanding and increasingly important role of administrative data.

The use of administrative information is seldom discretionary. In the formative years of any statistical agency, administrative data, such as trade data, have a central role in the statistical system. The data are cost effective to collect and compile, and the information on the documents is reasonably reliable. In the case of import and export data for agricultural products, no better source is likely to exist. For over 100 years, customs based data on exports and imports have provided valuable information regarding the markets for grains and livestock. For almost 70 years, administrative data on the grain trade have been integrated with census and survey statistics in a simple supply-disposition equation to assist in estimating the amount of grain required to meet domestic demand.

$$Demand = f(Production + (Imports - Exports) + Stocks - (Food + Feed + Seed + Waste + Dockage))$$

Administrative data are pervasive and the statistical system should recognise their existence and take advantage of their availability. All available and relevant data are handled best if they are confronted and reconciled within the context of the statistical system in terms of a balance sheet or input-output framework. If this is not the case, data users will eventually make use of the information to challenge

the system when, for example, estimated grain exports exceed the sum of domestic supplies, or when livestock estimates are insufficient to support reported slaughter levels, or consumption exceeds supply. The statistical system should be internally consistent, credible and make sense. Integrating the data, irrespective of the source, is an essential function of the statistician.

An inventory of a nation's statistical information, even in countries with no discernible statistical programs, will uncover a surprising amount of administrative information. In addition, as in the example of customs data, situations will occur where there will be more than one estimate. Agricultural imports are a good example. One estimate will be available from the Canada Customs data on imports and another from the sum of the exports to Canada tabulated from data provided by its obliging trading partners.

Making effective use of administrative data for statistical purposes is a challenge. Statistical system needs are not normally high on the priority list of most administrative data system developers. The objectives of an administrative data system are different from those of a statistical system, as are the definitions and concepts. Statisticians are seldom consulted in developing the capture, edit and imputation systems. Furthermore, the statistician is often given little information regarding the processing systems, little opportunity to follow-up issues with respondents, and little say in the reference period or the timing of the data's availability.

Theoretically, administrative data should be more reliable than survey and census data as a consequence of the regulatory systems put in place to review and verify the accuracy of the documents and enforce compliance. In reality, administrative data appear susceptible to the same sort of data collection and tabulation difficulties that plague census and survey data. Two separate administrative data systems collecting and compiling data using the same concept are not necessarily going to provide identical statistical information. The difference in the trade estimates between Canada and the United States is a classic example. Using information from Canadian customs documents, Canada produces estimates of imports from the United States and exports to the United States. Similarly, United States Customs using United States customs documents compiles estimates of imports from Canada and exports to Canada. Canadian estimates of imports from the United States do not match the United States estimates of exports to Canada, and Canadian exports to the United States do not match the United States estimates of imports from Canada. A ten-year joint effort by Canada and the United States to reconcile the differences ended with no satisfactory solution. It was a frustrating situation as both administrations were essentially measuring the same concept with similar methods and systems. As a common set of estimates was important for trade discussions, the agreed upon solution was to set the measurement issues aside. It was agreed that Statistics Canada would accept the United States estimates of imports from Canada as Canadian exports to the United States and the United States Department of Commerce would accept the Canadian estimates of imports from the United States as United States exports to Canada.

2. Agriculture Statistical Systems

Agriculture statistical systems are based on an integrated arrangement of statistical information from sample surveys, censuses and administrative sources. The system is a framework for linking information. Statistical systems are complex affairs that are usually constructed as a consequence of a long evolutionary process and often as part of an international cooperative effort. The similarities in requirements, and the desire for comparable statistics among nations, have meant that countries have traditionally cooperated in the development of statistical systems through organisations such as the International Statistics Institute (ISI dealt with a number of resolutions regarding agriculture statistics at

their 1895 meeting in Switzerland), the International Institute of Agriculture in Rome (established in 1908, it is the predecessor of the Food and Agriculture Organisation of the United Nations), the United Nations Statistical Office, the International Customs Union, the Organisation for Economic Cooperation and Development, and the United Nations Economic Commission for Europe. There are now many well-established concepts, conventions and statistical models that countries are encouraged to follow in developing their statistical systems. Nations that have established statistical systems have a vested interest in the current concepts, methods and conventions. This means that change is slow and considered, but not because they are unwilling to change. Nations with established systems have data users that want conceptual and measurement changes to have a minimal impact on the story told by the current and historical data and the international comparisons. Furthermore, the development and testing of new methods takes time.

In their formative years, statistical systems present a nation's socioeconomic profile by assembling the available administrative information. It consists, for example, of customs data on imports and exports, vital statistics on births, deaths, marriages, immigration and emigration, hospitalization and information on diseases and school enrollment. The administrative information is then complemented by information from a census of population, agriculture and industrial activity along with surveys of prices and employment, which provide further basic elements. Without exception, most nations have statistical systems that are composed of administrative, survey and census information, but the information is not always integrated.

3. Getting Value from Administrative Data

Many countries employ a mix of data from administrative, survey and census sources, integrating them together in their statistical system. The degree to which this is done, however, depends on a number of factors, the most important being: (1) the degree of maturity of the country's statistical system, (2) the quality and amount of information available from the administrative and regulatory programs, (3) well-trained and experienced staff, (4) funding, and (5) cooperation among government agencies.

Canada has moved to greatly extend the use and integration of agricultural administrative data over the past 70 years. This has been done for a number of reasons. First, administrative data reduce the need to ask farmers what they often perceive to be needless questions, thus helping to reduce respondent burden. Second, some types of farms are difficult to identify because their land base is very small. The types of farms that are the most difficult to identify are chicken, egg, pig and greenhouse farms because they use very little land, and also fruit and vegetable farms which tend to be concentrated in a few specific geographic areas. Administrative lists have helped address the problem. Third, experience has proven that farmer surveys tend to provide answers to financial questions that overstate expenses and understate sales. Fourth, most administrative data are of good quality, being compiled from documents completed in order to meet the requirements of an agricultural program or to ensure compliance with government regulations. Reviews and audits of the reported information by program administrators mean compliance and data quality, including coverage, are usually good.

4. The Integration of Administrative Data

4.1 Administrative Data to Supplement, Complement and Replace Survey Information

In the beginning, Canadian statisticians did not integrate administrative and survey data, although both sources were used to prepare the statistical publications of the times. Prior to Confederation in 1867, Canada undertook a series of censuses to record the areas under cultivation and the numbers of cattle

and sheep. This was done at frequent but irregular intervals beginning in 1667. The first of a series of regular decennial censuses of agriculture began in 1851. In the census of 1871, the first census after Confederation, both crop area and yields were collected for the principal field crops. In the census of 1891, horses, pigs and poultry questions were added to the livestock questions on cattle and sheep and, in the census of 1901, collection of crop information was extended to include all field crops [Godfrey 1985].

The settlement of western Canada and the rapid expansion in Canadian wheat production and grain exports were largely responsible for the *Census and Statistics Act* of 1905. The Act established a permanent national Census and Statistics Office within the Ministry of Agriculture charged with the responsibility to collect, abstract and tabulate certain classes of statistics, including those of agriculture. Responsibilities also included a provision for taking the future general decennial census of Canada and the population and agricultural censuses of the Prairie Provinces in the middle of each decade beginning in 1906.

Prior to 1906, there was no national information on crops and agriculture, with the exception of what was available from the decennial census of agriculture. At that time, the Maritime provinces, Ontario and each of the three Prairie provinces published more or less complete information on crops and agriculture; however, the efforts were not coordinated and the statistical methods varied widely among provinces. Quebec and British Columbia had no provincial agriculture statistics or crop reports. The only available information in those provinces was the decennial census of agriculture [Godfrey 1985].

In 1911, the national Census and Statistics Office was transferred to the Department of Trade and Commerce. In 1918, Canada passed the *Statistics Act*, which gave responsibility to what is now Statistics Canada to establish a comprehensive system of general statistics to address the needs and the demands of the times [Godfrey 1985].

One of the most pressing problems facing the new statistical agency was the confusion and uncertainty caused by the lack of cooperation between the national and provincial governments. The estimates published by the provinces differed from those of the new national statistical agency and the census of agriculture. The confusion resulting from more than one official government estimate and the duplication of federal and provincial efforts was recognised as unsatisfactory by all governments. The *Statistics Act* of 1918 gave the Minister responsible for Statistics Canada authorization to enter negotiations with the provincial governments with a view to securing a better measure of cooperation. Since 1918 and the first cooperative efforts, the publications of the national and provincial governments have contained identical figures. Federal-provincial cooperation remains an important element of the *Statistics Act* and the underlying principal behind Statistics Canada's annual federal-provincial meetings on agriculture statistics. The detailed collection and data-sharing arrangements, however, have changed over time and continue to be somewhat flexible in character in order to address provincial as well as federal requirements.

Until 1930, agriculture statistics were largely limited to crop areas and yields, livestock numbers, and the production, prices and value of crops, livestock and livestock products. Annual statistics on the visible supplies of grain and other agricultural commodities, exports and imports and the slaughter and marketing of livestock and animal products were tabulated from administrative data. It was not until 1930 that efforts were made to integrate this administrative data with the annual agriculture survey data in an input-output (supply-disposition) framework to provide final crop and livestock estimates, thus ensuring that the historical series was consistent with the disposition data on marketings and exports.

Since 1930, Canada has made extensive use of administrative data on agriculture product imports and exports, domestic marketings and livestock slaughter to improve the quality and revise the official estimates. For the 50 years between 1930 and 1970, administrative data were used mainly as a complement to survey and census data in the context of supply and disposition analysis. During that period, active efforts were made to identify and employ new and emerging data sources. In 1935, the establishment of the Canadian Wheat Board greatly improved the quality and timeliness of administrative data on wheat, oats and barley marketings and the supply and disposition estimates of grains. At about the same time, what is now called the Canadian Grain Commission began to publish information on exports, a consequence of its inspection and regulatory responsibilities. Increasingly rigorous inspection and grading regulations have also steadily improved the quality and coverage of administrative data on livestock slaughter, egg production, dairy product production and fruit and vegetable marketings.

The improvements in the supply and disposition tables that began in the 1930s, with the incorporation of administrative data on marketings and trade, resulted in the publication of a number of new value-added statistics on food supplies. These data proved invaluable to Canada during the Second World War in estimating domestic food requirements and the availability of products for export to assist with the war effort. These estimates of domestic food supplies are still compiled in Canada. Their value to the average Canadian is no longer from a food security perspective but in relation to trends in the national diet and whether, as a nation, Canadians are eating a more healthy diet with fewer fats and oils and more fresh products and fibre. Many people and countries in the world, however, are still concerned with basic food supplies and food security. International agencies, such as the Food and Agriculture Organisation, continue to tabulate and publish food supplies by commodity and encourage all countries to establish food supply estimation procedures as an element of the national statistics system.

4.2 The Survey Process and Administrative Data

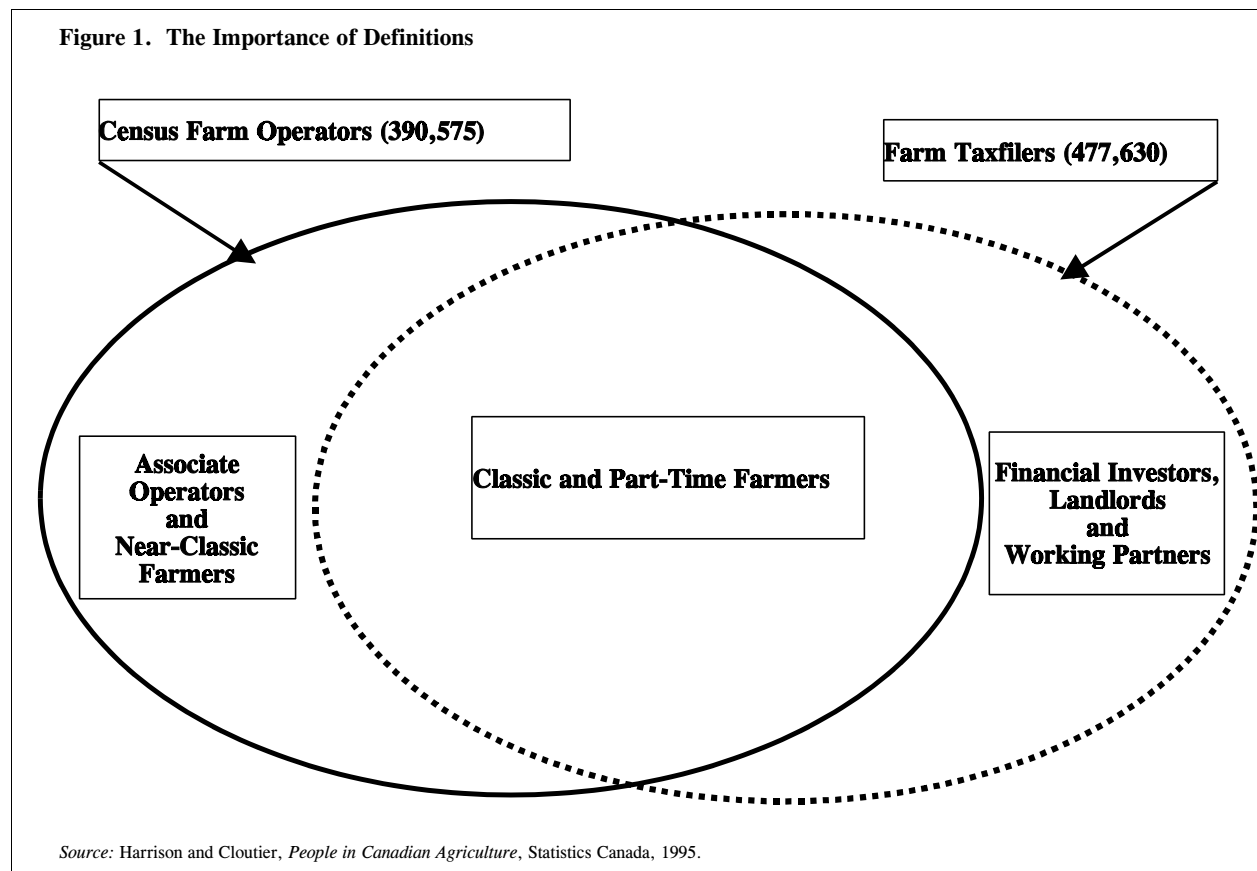
Administrative data have numerous uses in the survey process. Administrative data can be used for: (1) frame creation and list maintenance, (2) stratification and sample selection, (3) assisting with editing, imputation and estimation, and (4) comparative analysis which is useful in validating survey information.

The mid-1960s in Canada saw the emergence of supply management for chickens, eggs and milk in response to Canadian agriculture's successful creation of an overabundance of products. The newly available marketing board statistics on production complemented the administrative data already available on marketings. The supply management boards also maintained producer lists as part of their regulatory function. This new information on producers raised the possibility that the information could be used to improve the Farm Register, the listing of farmers used to draw sample surveys. Although marketing board lists did not include all farmers, the annual, up-to-date list of chicken, egg and dairy farmers offered the potential for significant improvements to list maintenance, because until then the Farm Register was almost totally dependent on the quinquennial census of agriculture for updates.

The period following 1970 also marked the beginning of a more aggressive effort to use administrative data in order to: (1) improve the quality of information traditionally provided by surveys, (2) supplement survey data to reduce response burden, and (3) maintain the Farm Register. Administrative data also continued to be used to improve the quality of official statistics in the framework of supply and disposition analysis.

The use of administrative data to improve the quality of information, previously available only from the census of agriculture and surveys, uncovered new and numerous challenges. The most ambitious initiative was the Whole Farm Data Project, a project established in the early 1990s to assemble and integrate information to provide a statistical image of “whole farms”, namely their crop, livestock and financial information. The project involved reconciling survey commodity data and financial information from administrative sources to produce a view of the “whole farm”. In addition to the well accepted understanding that administrative data were not always as timely as survey information, the new uses of the data highlighted a number of other differences. Administrative data were not always conceptually the same as survey data and, even when they were identical, the reporting or compliance with the particular program or regulation often resulted in quite unexpected statistics from producer lists [Pursey 1981].

Statistics Canada has for some time been using tax records to provide estimates of farm expenses in an effort to reduce response burden [Gellner 1972]. The 1991 Census of Agriculture reported 280,043 farms and 390,575 census farm operators in comparison to the 477,630 individuals who filed 1991 farm tax returns. Figure 1 provides a graphic explanation of some of the complexity. A cautionary note for the diagram is that it is not to scale as efforts to estimate the number of individuals in each of the overlap groups were not successful. Consequently, it is not a good indicator of the real size of each group or the actual amount of overlap.



Understanding the underlying concepts and definitions of each database is critical to the success of data integration initiatives. For example, there is no limit to the number of tax returns that may be filed on a farm operation nor does a taxfiler have to be a farmer to report income from farming. Landlords, investors or the spouse and children of farmers can all be farm taxfilers. Statistical systems are a consequence of evolution, and the level of data integration that is achieved is often an indication of the degree of system development. To add to the complexity described in Figure 1, it appears that a substantial number of census farmers do not file farm tax returns. These farmers are not evading or avoiding taxes; they did complete a tax return, but it was not a farm tax return.

Efforts have been underway for some time to maintain the Farm Register with information from producer marketing boards and farmer associations. With the lists being incomplete, however, progress has been modest. Marketing board lists only include producers of the product for which the board has jurisdiction. In the case of producer associations, not all producers are members. In addition, not all boards and associations have been willing to share their information. Some work was undertaken in an attempt to use information from farm magazine and newspaper subscription lists for list maintenance, but it was not particularly successful, as there were too many non-farm subscribers on the lists for them to be useful.

Despite the effort, the work to discover how administrative data can be used for list frame maintenance is worthwhile, as the potential returns are significant. Recently filed tax records are being evaluated as a means to maintain names and addresses on the Farm Register [Denis and Whitridge 1977]. The Canadian census of agriculture has traditionally been collected at the same time as the census of population. Data collection procedures in the past have instructed census enumerators to leave a census of agriculture questionnaire with every household reporting a farm operation. This resulted in efficient data collection, good coverage of farms and a good farmer list from the census. It is possible that the census of agriculture will be conducted entirely by mail in the year 2006. This will mean that, instead of the census of agriculture being the basis of a list frame, it will require the names and addresses of farmers for a mailing list.

5. Observations and Conclusions

The majority of the statistical programs in both developed and developing countries tend to operate with modest resources in an environment of scarce and sometimes reduced government funding. Nevertheless, the statistical programs of the developed countries are often found on the leading edge of new frontiers where they pioneer methods such as new applications for administrative data. The statistical systems of developed countries are established but not static. The social and economic fabric of countries is forever changing and, as a consequence, the statistical systems are subject to a continuous process of adjustment and modification in an effort to address emerging needs and priorities.

Developing nations face different circumstances. They may often have the good fortune of being able to model and implement the concepts and methods of statistical systems that have already been established. They must also deal with the complexities associated with the need to accomplish the development process in a very short time and simultaneously assimilate and keep pace with the rapid advances in informatics and data processing technology.

Most countries appear to have gone through an evolutionary process in establishing their agriculture statistical system. However, the development route that is followed is highly dependent on: (1) the amount of resources available for the program, (2) the availability of experienced statisticians and

related professionals to develop and maintain the system, and (3) the toolbox statisticians are able to use. Because of a lack of resources or experienced staff, many statisticians and statistical agencies do not take full advantage of the range of resources in the toolbox which includes administrative data, area frame surveys, list frame surveys, multi-frame surveys and censuses of agriculture. They may also face additional difficulties such as a lack of cartography, lists or willingness of other government authorities to share administrative data.

The integration of survey, census and administrative data takes a substantial effort. Development requires well-trained and experienced staff, funding, access to the necessary statistical methodology and cooperation among government agencies. Statistical agencies exist in order to provide governments with accurate, relevant and reliable statistical information, but their long-term survival depends on their ability to evolve and adapt.

References

Denis, J. and Whitridge, P. (1997), "Agriculture Statistics – Future Developments", Presented to the Advisory Committee on Statistical Methods, Statistics Canada, Ottawa, Canada.

Gellner, J.A. (1972), "Farmers Income Tax Returns as a Source of Agricultural Statistics," Summary Report of the 23rd Federal-Provincial Conference on Agricultural Statistics, Statistics Canada, Ottawa, Canada.

Godfrey, E.A. (1985), *Progress in the Collection of Annual Agricultural Statistics and of Crop Reports During Growth in the Dominion of Canada, 1908-1935*, Methodology Paper No. 9, Catalogue No. 21-520, Agriculture Division, Statistics Canada, Ottawa, Canada.

Harrison, R. and Cloutier, S. (1995), *People in Canadian Agriculture*, Catalogue No. 21-523, Agriculture Division, Statistics Canada, Ottawa, Canada.

Purseley, S. (1981), "Farm Operating Expense Estimates in British Columbia, 1977-1978, A Study of the Feasibility of Substituting Tax Data for Survey Data," Unpublished, Agriculture Division, Statistics Canada, Ottawa, Canada.