Integrating the Canadian Annual Survey of Manufactures into the Unified Enterprise Survey: Challenges and Issues

Isabelle Marchand, Emmanuel Benhin and José Gaudet

Statistics Canada, 120 Parkdale, Ottawa (Ontario), Canada, K1A 0T6
isabelle.marchand@statcan.ca, emmanuel.benhin@statcan.ca, jose.gaudet@statcan.ca

1. Introduction

The Annual Survey of Manufactures (ASM) is one of the largest of all the business surveys conducted by Statistics Canada. It is one of the primary contributors to the calculation of the Canadian Gross Domestic Product, representing almost 20% of the economic production. The ASM covers all establishments belonging to the manufacturing sector and the logging industry. It collects various financial data such as employment, wages, total cost of raw materials (referred here as input total), total sales of manufactured products (referred here as output total) and inventories. It also collects detailed commodity data such as the cost of each raw material, the sale of each manufactured product, and the quantity of each raw material or manufactured product. As part of a strategic streamlining initiative that aimed to reduce survey costs and resources while maintaining a certain level of data quality, the ASM conducted some major changes for the reference year (RY) 2004. One of the objectives of these changes was to maximize the use of fiscal data. These changes affected many components of the survey process such as the sample design, questionnaire contents, methodology and processing systems.

One of the major changes was to further integrate the ASM into the Unified Enterprise Survey (UES). These two surveys have shared a few survey processing components since RY2000. The UES is an annual business survey that serves as the vehicle for producing annual estimates for many industries at a variety of geographic levels. The different annual surveys that are part of the UES are integrated into one platform where a single frame, similar methodology and processing systems are used. Another important objective of the UES is to reduce response burden of Canadian businesses by increasing the use of administrative data (Gauthier, 2005).

The paper is organized as follows: Section 2 provides an overview of the ASM and the UES programs. In Section 3, we summarize, in general terms, the benefits and limitations of survey integration. We provide some key challenges of the ASM integration process. We also discuss the integration processes for each of the ASM survey steps and the associated challenges. Finally, we give some advantages and disadvantages of integration of the ASM into the UES.

2. An overview of the ASM and the UES programs

2.1. Overview of the ASM

The Annual Survey of Manufactures covers all establishments belonging to the manufacturing sector and the logging industry identified using the Northern American Industrial Classification System (NAICS). These sectors are divided into 23 industries at the NAICS3 level (NAICS coding using the first three digits). To reduce cost and response burden, exclusion thresholds were used to divide the target population into a take-none portion (units below the exclusion thresholds) and a survey portion (units above the exclusion thresholds). From RY2000 to RY2003, the ASM used a stratified two-phase sample design without replacement. Stratification of the survey portion was done by industry (NAICS3), geographic region (provinces and territories) and by size of revenue. The first phase identified units that received long form questionnaires (to collect detailed financial and commodity data). The second phase identified units that received short form questionnaires (to collect a summary of financial data). In RY2003, there were no longer short form questionnaires and so the second phase identified units that were directly replaced by fiscal data. Estimates of totals were produced at the NAICS6 by geographic levels for the survey portion based on census-like financial data and commodity data. These census-like data were created by imputing for all the out-of-sample units in the survey portion. This was accomplished by various imputation and modelling techniques using a combination of fiscal and respondents' data. Estimates from the take-none portion were produced for a subset of the financial variables using only fiscal data.

For RY2004, one of the objectives of the ASM was to maximize the use of fiscal data through changes to the sample design, questionnaire contents, methodology and processing systems. The use of fiscal data contributes to decreasing the cost and

response burden. The ASM has always used fiscal data to produce estimates but the type of data was expanded for RY2004. A new questionnaire was developed for RY2004 that collects financial data in accordance with the Chart of Accounts specifications (Statistics Canada standards for collecting financial data). With this new questionnaire, 10% of the ASM financial variables were directly replaced by fiscal data. The use of fiscal data was also expanded to the whole population and so there was no need for a second phase sampling for the RY2004 sample design. Financial variables were estimated using the census-like financial data. While developing the new sample design, consultation with the System of National Accounts (SNA), the main user of ASM data, led to the decision that there was no longer a need for a complete set of micro-data (census-like) for commodities. Therefore, commodities were estimated via a weighting scheme (Gaudet and Marchand, 2005).

2.2. Overview of the UES

The UES program covers many sectors of Canadian economy: distributive trade, services, transportation and agriculture involving more than 25 surveys. The SNA uses the data from the UES to produce and publish its statistics such as the Canadian Gross Domestic Product. For RY2004, the target population of the UES was also split into survey portion and take-none portion. The survey portion used a stratified two-phase sample design. Stratification was by industry (using NAICS codes), geographic region (province and territories) and size of revenue. The first phase identified the global sample units and the second phase identified those units that received fiscal data. Principal financial statistics are produced using weighted estimation for the survey portion. Estimates from the take-none portion are based on only fiscal data.

3. Integration of the ASM into the UES

3.1 Survey integration

This section outlines some advantages and disadvantages of survey integration, in a general context. There are many advantages when surveys are integrated although one may lose some flexibility with respect to individual surveys. However, the advantages of decreasing cost and improving performance and data quality outweigh the disadvantages.

In general, the advantages of survey integration are:

- Reducing cost by centralizing collection, processing and management.
- Reducing response burden by synchronizing and harmonizing collection periods, questionnaire contents as well as having common respondent identifiers.
- Increasing consistency by using common statistics, concepts, definitions and methods. This also makes it easier to share information among partners since information can be summarized in the same way and fairly quickly.
- Simplifying and increasing the efficiency of processing. Having a single processing system is easier to maintain than many different systems that are doing roughly the same thing.
- Sharing best practices and solutions among surveys that have similar constraints and problems.
- Innovation, development and research may be pursued to benefit several different survey programs at the same time.
- Improving corporate planning and coordination

Disadvantages of survey integration include the following:

- Not having a survey design and a processing system that are tailored to the specific requirements of different surveys so that changes in methodology can be handled within time and budget constraints.
- Managing communication, decision making and processing within a large survey program can be challenging.
- Losing flexibility for each individual survey program as well as personalized services.
- Although simplicity is targeted, complexity may be the result from changes made to address specific needs of individual survey programs.

3.2. Challenges from the ASM integration into the UES

The ASM and the UES have been sharing some common survey processing components since RY2000. The two surveys share the same frame, collect common financial variables, use the same collection tools, and the same processing systems for the in-sample portion of the ASM financial data. The in-sample portion of the ASM financial data is processed by the same team responsible for the UES financial data. In addition, the two surveys follow the same schedule for each survey step, since both are required to publish the financial estimates within a 15-month period. However, many aspects of these two surveys

remain different. The ASM and the UES have different sample designs, different methodologies, different processing systems for the out-of-sample financial data and the commodity data. They were managed independently by different teams of methodologists. For RY2004, the following were targeted for further integration: the ASM and the UES are to be managed by one team. Methodologies and processing systems are to be harmonized when desirable and feasible. The transition to a more integrated ASM for RY2004 posed a few challenges, namely:

#1: Definition of integration. Integration can be interpreted in different ways. Since RY2000 the ASM has been sharing some common components with the UES. For some, this meant that the ASM was already "integrated". However, there was no general agreement about the desired extent of integration of the ASM into the UES. This complicated communication, planning and decision making.

#2: Integrate a survey with a different sample design. Although there are some differences among surveys that are part of the UES, the ASM remains the only one with a completely different sample design. The RY2004 sample design was primarily developed to meet specific user's requirement and also to maximize the use of fiscal data. The differences in concepts, definitions and methods made sharing of information particularly difficult and as such the integration process more challenging.

#3: Processing commodity data. The ASM collects two types of data: financial and commodity data. For the ASM RY2004, the commodity input totals and output totals breakdown into more than 1,400 types of detailed commodity codes (codes for each raw material and manufactured product). The amount of information to process was significant and for this reason commodity data have always been processed separately from the financial data. There is one survey in UES that collects commodity data, but the level of details is much smaller and therefore, in this particular case, the commodity data are handled like the financial data. As a result of the ASM integration, new modules needed to be built in the existing UES processing systems, and eventually maintained by one team. In addition, staff working on UES needed to develop a new expertise to handle the expanded processing systems.

#4: Changing the work environment: from a single survey to a large survey program. A large survey program cannot be managed the same way as a single survey program. The UES is managed through a matrix management approach. In the UES, a division which supports a survey program (services, trades for example) is responsible for coordinating questionnaires, collection processes and data management operations. Methodologists provide methodological support to all the partners of the UES (Gauthier, 2005). Methodologists responsible for the ASM are expected to be part of the UES methodologists' team, where sampling, collection, edit and imputation (E&I) and estimation are handled by separate units. The two working environments are different and so integration of these environments meant an increasing challenge on the work of the ASM methodologists.

3.3 Integration process of the ASM into the UES for RY2004

In this section, we describe for each survey step, how the ASM RY2004 was further integrated into the UES with respect to concepts, definitions, methodologies and processing systems given the constraints on time and resources.

3.3.1 Frame

The ASM and the UES continue to use the same frame; the Business Register (BR), a database from Statistics Canada that lists all the non-incorporated (T1) and incorporated businesses (T2) known to be in operation in Canada. On the BR, the statistical structure of a business consists of four levels of statistical entities: enterprise, company, establishment and location. One enterprise can have one or many companies, one company can have one or many establishments, etc. For many businesses, the four levels coincide. These are referred to as simple businesses. All other businesses are referred to as complex businesses.

3.3.2 Sampling

For all aspects of the ASM sampling process, the possibility of using the UES methods was evaluated and implemented when deemed desirable. The sampling unit for the ASM is the establishment. In the ASM, all establishments that belong to a complex business were selected with certainty. Currently, fiscal data are available only for simple businesses. The fiscal data and the respondents' data were used to create a census-like financial data file. For the UES, the sampling unit consists of all

establishments in the same enterprise within an industry and a geographic region. Network sampling was used to ensure that when an establishment which was part of a complex business was selected, then all related establishments were also selected. The two methods lead to the same result in the sense that all establishments within a complex business were selected with certainty. However, due to the constraints on time and resources, the ASM retained its sampling methodology.

Determination of targeted coefficients of variation (CVs) for the sample sizes in the ASM was done by having the same CV for each combination of industry by geographic levels. As part of the integration process, the ASM method, the UES method, and a third method were evaluated. In the UES method, CVs at the national level for each industry and each geographic region were first determined. Then CVs at all industry by geographic levels were derived according to the significance of the industry while aiming at similar levels of precision across all geographic regions. For the third approach, the same CVs for each geographic region and for each industry as in the UES method were used. However, a raking ratio approach was used to derive targeted CVs for all industry by geographic levels. The results of this study were presented to the subject matter officers and the SNA team, the primary user of the ASM data. The UES method was adopted because it met specific requirements for the SNA. For stratification and allocation of the sample, the ASM and the UES use the Lavallée-Hidiroglou algorithm (Lavallée and Hidiroglou, 1988). This algorithm stratifies asymmetric populations in such a way that, for a fixed level of precision (CV), it minimizes the global sample size.

The processing systems for the sample selection were different for the two surveys. The ASM used Statistics Canada's Generalised Sampling System (GSAM) and the UES used Micro-strat, a sample selection system developed by the National Institute of Statistics and Economic Studies in France. For the integration process, it was decided that the ASM continue to use GSAM since getting support was easier given the fact that it is a Statistics Canada system. The UES will be adopting GSAM for RY2005.

3.3.3 Collection and score function

The two surveys used the same collection systems. Collection was managed by the same team as operations. To manage collection efficiently, the two surveys used a score function to rank units and prioritise follow-up calls. The methodology was very different. The ASM used a score function where survey units were prioritised based on their contribution to output totals and to the distribution of commodity type. A unit score was updated based on the information collected from respondents. The number of times a respondent was contacted was also a parameter in the ASM score function. The UES score function was simpler and used information solely from the survey frame. It prioritised units by following respondents with the largest revenue total. However, this variable was not updated as data from the respondent became available. Due to time and resources constraints, the ASM score function was not reviewed prior to implementation for RY2004. We plan to review these methods in the near future.

3.3.4 Edit and Imputation

Prior to RY004, processing of the E&I for the ASM and the UES were done by different teams and based on different platforms and processing systems. Different methods were used for the financial in-sample, financial out-of-sample and commodity components. These three components were processed sequentially in the order above. The different processing systems were costly to maintain, as well as complex to follow. The transfer of data among these components was a challenge in itself. As mentioned in Subsection 3.2, the financial in-sample portion already uses the UES platform. The integration of the other components of the ASM processing systems into the UES was reviewed with the goal of having a single processing system and managed by a single team. For edit and imputation, the UES uses Banff, a generalized system from Statistics Canada that is based on SAS (Statistical Analysis System). Banff covers different E&I modules such as historical, donor and ratio that may be used by various surveys.

Financial variables Before RY2004, there were two separate ASM processing systems, one for the in-sample portion, and another one for the out-of-sample portion. The in-sample portion was processed using the UES platform. The system was managed by the UES team responsible for the E&I processing. The methodology to process the data was as follows: to use tax data when available, then historical imputation if historical data were available, then donor imputation if no historical data were available and finally, massive imputation. The out-of-sample portion was imputed by the ASM methodology team, using in-house SAS programs. For imputation, various models were derived at different levels of industry and geographic regions using a combination of respondents' data and fiscal data.

For RY2004, the ASM E&I for the in-sample and out-of-sample will be processed by one system, using the current ASM in-sample system. The new methodology will not use modelling techniques and will rely less on historical imputation. As well, donor imputation will no longer be used since this method was not yielding satisfactory results. The new E&I methodology will first consist of using tax data when available, then historical imputation (using only respondents' data), then ratio imputation and finally, massive imputation. This is similar to the methods used by the UES. The UES does not use a census-like approach and only has to process the in-sample portion. In general, the UES uses the following imputation methods in this order: use tax data to impute when available, ratio imputation, historical, donor and massive imputation.

Commodity variables For ASM, the quantity of commodity information to be processed was quite large. This is why these data were processed independently from the financial variables. Prior to RY2004, the commodity data were processed by the ASM methodologists using in-house SAS programs and the Standard Economic Processing System (StEPS), a SAS-based software that originated from the United States Census Bureau. The methodology was essentially as follows: historical imputation was used first, then donor imputation if there were no historical data. All the commodities were then pro-rated to equal the input totals or output totals, two financial variables already processed. For RY2004, a new module was created in the UES environment for commodities. The processing system is Banff and the methodology remains essentially the same. The main difference was that only in-sample commodity data were imputed.

3.3.5 Estimation

The ASM and the UES produce estimates of totals for financial variables (e.g., employment, wages and inventories) and commodity variables (e.g., cost of a raw material, sale of a manufactured product and quantity of a manufactured product).

Financial variables Prior to RY2004, ASM estimates of totals for the survey portion were produced by simple aggregation based on census-like financial data and census-like commodity data. This was possible since mass imputation was performed in order to have complete population files. The UES on the other hand produces estimates of totals using sampling weights. Different processing systems were required for the ASM and the UES.

For the take-none portion, the ASM and the UES produce estimates of totals for a subset of the financial variables. The estimates were based on contributions from the T1 and the T2 tax data. Estimates from the T1 were based on sampling weights and the T2 estimates were obtained by simple aggregation. The same processing systems were used for the ASM and the UES.

For reference year 2004, the ASM will continue to produce estimates of totals for the financial variables using simple aggregation for the survey portion. This method is different from the weighted estimation approach used in the UES. The processing system required to implement this component of the ASM method within the UES necessitated minor expansions to the existing UES system. This was handled primarily at the E&I processing stage. For the take-none portion, the ASM will continue to use the existing ASM/UES estimation method and processing system with minimum modifications.

Commodity variables The ASM estimation of totals for commodity variables for RY2004 changed from simple aggregation of census-like commodity data to sample-based estimates of totals for the survey portion. A calibration method of estimation was adopted. This method adjusts the sampling weights so that estimates of commodity totals for the input totals and output totals are consistent with the corresponding estimates of census-like financial totals within each NAICS6 by geographic level. The integration of this component of the ASM estimation method within the UES processing systems required the development of new UES components. The new components required the integration of the Generalized Estimation System (GES), associated Pre-GES and Post-GES systems into the existing UES processing system. The GES is an estimation system developed in Statistics Canada to process many well known estimation methods used in several Statistics Canada surveys.

3.4 Advantages and disadvantages of integration of the ASM into the UES

Subsection 3.1 listed advantages and disadvantages of survey integration in a general context, mainly from the perspective of the survey that was targeted to be integrated. However, when two surveys are integrated, both are impacted in many different ways. This section presents the advantages and disadvantages that have been noted for both the ASM and the UES programs.

Table 1. Advantages and disadvantages of the ASM integration

Advantages	Disadvantages
• Reducing general cost • Everybody working on ASM from any expertise area is learning about other similar survey procedures, methodology and concerns • A single and simplified processing system is now used and will be easier to maintain and understand • Decrease in processing time	• Many commitments to meetings, committees and working groups that take time and are not always relevant to the integrating survey • It is not always clear to subject matter who can answer methodology questions or solve a problem in this new large environment. • Methodologists who used to work on ASM lose access to broader understanding of the overall survey process by becoming members of a specialized unit for a given survey step • Problems in processing one UES survey can now affect the ASM

Table 2. Advantages and disadvantages of survey integration for the UES

Advantages	Disadvantages
• New modules have been put into production and could potentially be used by other surveys in the future: o Processing system for commodity data o Use of GES for estimation o Use of a calibration method for estimation Opportunity to learn from a different survey's approach and methods: • A different way of using fiscal data • A more sophisticated score function to prioritize follow-up calls	• At the beginning, increased learning curves as all members need to learn the details of the new survey. • Increased challenge in communicating to partners two different sample designs within the integrated survey system; one for the ASM and another for the non-ASM. • Increased workload on every aspect of the integrated survey: production, management, development and maintenance. • Increased processing time • ASM processing can now affect other surveys

4. Conclusion

In this paper, we have provided some background to the ASM and the UES programs. We have outlined some challenges of the ASM integration into the UES. Some of the main challenges were related to communication issues. Due to the competing needs, the scope of the integration project became dynamic. The constraints on time and resources made this dynamic project even more challenging. We also discussed in this paper, the integration processes in relation to harmonizing concepts, definitions, issues of methodology and processing systems. We presented some benefits and limitations of integrating the ASM into the UES. For the ASM, the gains are a general reduction in cost of survey, a decrease in processing time and a simplified processing system. This processing system will also be easier to maintain. For the UES, introduction of new methodologies and additional processing systems may be used by other UES surveys.

For future work, we propose to first review the effects of the integration of the ASM into the UES on several components of the ASM survey process and study further harmonization of concepts and methodologies. Another important future work for the ASM is the development of quality indicator tools for estimates for the financial variables. This may involve the development of variance estimation that accounts for imputation. This is an area that has been investigated by the UES and in the process of implementation. In this respect the ASM could benefit from the knowledge of the UES.

References

Gauthier, S. (2005). De 7 à 26 enquêtes – complexité grandissante de la gestion de l'échantillonnage pour l'enquête unifiée auprès des entreprises. *Colloque francophone sur les sondages 2005*. Québec, Canada.

Gaudet, J. et Marchand, I. (2005). Le remaniement du plan d'échantillonnage de l'Enquête annuelle sur les manufactures. *Colloque francophone sur les sondages 2005*. Québec, Canada.

Lavallée, P. and Hidiroglou, M. A. (1988). On the Stratification of skewed populations. *Survey Methodology.* 14, 33-43.