

## **2011 European Establishment Statistics Workshop – EESW11**

### **Work session on Integrated Business Statistics Program**

(12-14 September 2011, Neuchâtel, Switzerland)

#### **DEVELOPMENT OF AN INTEGRATED BUSINESS STATISTICS PROGRAM**

Prepared by Jean-François Carbonneau, Statistics Canada

#### **I. Introduction**

1. In 2010, Statistics Canada introduced the Corporate Business Architecture Initiative (CBA) to conduct a comprehensive review and revision of Statistics Canada's business architecture in order to harvest efficiency on ongoing operating costs of 5% within 5 years, enhance quality assurance through implementation of more robust systems and processes and improved responsiveness in delivery of new statistical programs. Combined with the challenge of improving relevance while maintaining quality, this initiative will force Statistics Canada to achieve a much greater efficiency than at any time in the past. It is within this context that StatCan is looking at significantly centralizing and standardizing the process involved in the production of statistical output. The proposal to develop an Integrated Business Statistics Program (IBSP) represents a way of achieving these objectives for business statistics. Initiated in 2010, the IBSP is to be based on the existing model the Unified Enterprise Statistics program (UES) which had been conceptualized in 1996 and implemented through different phases up until 2008.

2. Generalized approach to survey processing cannot satisfy the particular needs of all existing programs. Rather, it will be necessary for programs that they adjust up to a certain degree to fit into the generalized model. This is the only way to ensure that with a limited number of resources we achieve the highest degree of relevance while maintaining data quality. Currently the Unified Enterprise Surveys Program (UES) is applied to 59 annual business surveys and covers the sample and questionnaire design, edits and imputation, allocation and estimation processes.

3. By 2016, close to 130 annual and sub-annual surveys in ten different programs will be integrated into the IBSP framework. The surveys under the umbrella of the IBSP will use the Statistics Canada's Business Register as a common frame. They will adopt electronic data collection as the principal mode of collection, focus on a common sampling approach, use tax data universe for the estimation of financial information, apply a common editing strategy for automated and manual editing, establish an earlier collection cut-off and expand on the use of data warehouses. All processing methodologies will be driven by a common metadata

framework and they will share common tools to analyse, edit and correct data.

## **II. Questionnaire design and content determination**

4. The IBSP questionnaire will follow a modular approach and contain one core financial module and various characteristics modules dedicated to collecting information for specific industries. Two types of questionnaires are envisioned. The first will be dedicated to collecting data at the enterprise or establishment level from the very complex businesses for which financial data cannot be obtained from administrative data. It will be a collection vehicle containing all questions on financial as well as characteristics information. For very complex businesses taken care by Enterprise Portfolio Managers, they will be able to customize the questionnaire in order to capture data for the entire enterprise. The module dedicated to the capture of characteristic information will be customized by industry or activity. The second questionnaire would be a simplified version of the first, dedicated to the collection of industry-specific commodity or characteristics data at the establishment level. This questionnaire will be targeted to the majority of businesses that have simple structures with their financial information coming from administrative tax data and made up of a number of modules, each dedicated to collecting a specific type of characteristic (e.g. input or output commodities, origin and destination of shipments, qualitative type of information).

5. In order to determine questionnaire content, consultations with the various subject matter areas have been held. Crucial to this exercise is the input of the System of National Accounts. Standardized questionnaire modules will adhere to the North American Industry Classification System (NAICS), North American Product Classification System (NAPCS) and the Tax data Chart of Accounts. This will ensure coherence across the board and it will eliminate the manual mapping between tax and survey data that occurs presently.

## **III. Current UES Process Vs Proposed IBSP Process**

6. UES collection consists of the pre-contact, mail out of questionnaires and the non-response and edits follow-up activities. Score functions (SF) based on the contribution of a unit to the total value of a key financial variable, are dynamically applied to minimize costs. With the exception of six Services Industry surveys, all UES surveys are currently collected on an annual basis. The six exceptions have an “on and off” strategy whereby every second year only certain modules of the questionnaire are sent out. As for 300 large and complex enterprises, collection is performed by Enterprise Portfolio Managers using paper and EXCEL questionnaires which are sent and received via e-file transfer. Typically, the mail out activities occurs during the January to March period, while edit and non-response follow-up extend up until the end of October.

The UES follows a standard linear survey process where collection activities (including failed collection edits follow-ups) are performed first, followed by processing tasks, such as Edits, Imputations (E&I), Allocation and finally the generation of estimates. After each of these steps, data validation and manual editing is performed by the subject matter analysts. In order to improve the timeliness of the survey process, integration must start at the collection step.

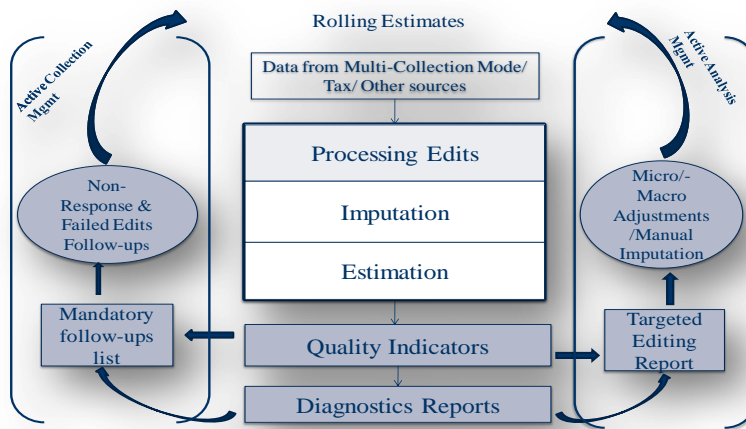
7. The IBSP proposes to follow an iterative process that will aim at optimizing the editing work done by both collection services (failed collection edits follow-ups) and subject matter divisions. Quality Indicators and decision rules will be provided to better target units that need a follow-up and assess the relevance of manually editing a record. The efforts devoted to editing are expected to decline considerably. The implementation of a new processing model supported by a Common Editing Strategy (CES) is central to the IBSP project. Under this iterative approach, collection, processing and micro/macro editing activities will be done in parallel and not in a sequential way. The objectives of the CES are to contribute to the reduction of operational costs via the harmonization of the editing methods and tools, the expansion of automation in the editing activities and also the reduction of respondent follow-up activities and manual interventions on non-influential units. It is also expected that this common editing framework will be beneficial to data quality with improved timeliness of survey results. The accuracy and the coherence of statistical information are other elements of the data quality that should be positively affected by re-directing subject matter editing efforts toward analytical activities.

8. Recent studies on the volume and impact of manual interventions on the Unified Enterprise Survey (UES) and interviews with analysts from all programs under the umbrella of the IBSP on their data editing practices emphasize the need for rationalization and harmonization of these activities. Each survey area has, over the years, developed its own set of tools and practices for performing data adjustments. With the integration of new business programs into the IBSP framework, a revamped processing model driven by an efficient Common Edit Strategy (CES) will be put in place. No manual intervention will be possible either before or after E&I during an iteration as it is the case in the current UES processing framework. The edits, imputation, allocation and estimation functions are performed on the data without any manual intervention. Data is reviewed only after the estimates are available in the analytical tool. The estimates will be assessed based on macro quality indicators after each iteration. In subsequent iterations, corrections and manual adjustments on influential records are integrated along with the new survey and available administrative data.

9. Based on both the macro quality indicators (calculated by estimation domains) and the micro record impact score, data editing resources would be assigned in priority on units with a

high impact in their domains of estimation. Domains where a quality improvement is required to meet the targets will be prioritized. Subject matter areas will be able to review influential records in relation to the estimates and their quality (Active Analysis Management). Collection Services will conduct follow-ups based on a priority list of units based on the current survey results (Active Collection Management). Cost reduction is the main objective of this model. It is hoped to achieve this through the reduction of manual interventions, optimization of follow-up activities from the CES. Timeliness should also be improved since estimates would have already been reviewed several times and most influential records validated and edited by the time the last iteration is run.

Figure 1 – Rolling Estimates and Common Editing Strategy



#### IV. Generalized systems and Analytical tools

10. A key component of this initiative is the development and mandatory use of shared and generic corporate services and systems for processing, analysing disseminating and storing statistical information for business programs. The objectives is to limit the diversity of tools developed and maintained and ensuring the harmonization of concepts and training activities. The generalized systems are based on standardized methods which we will be available to survey managers. All systems must be metadata driven and repositories for the metadata must be developed. It is also important that a common platform be used by the majority of systems, to avoid the need for data transformation steps.

#### V. Conclusion

11. The system shown in Figure 1 will be developed in phases. A parallel run will be conducted on the 2011 production data. The purpose of this exercise will be to put to test the new design of the Rolling Estimates and Common Editing Strategy. More specifically we will assess the newly introduced functionalities such as 1)Monitoring collection and analysis progress and prioritizing collection and analysis efforts using Quality Indicators, 2) Replacing

financial Survey data with Tax data for simple units, 3) Impact of an iterative process on timeliness and data quality, 4) the use of administrative data in replacement of survey information related to the allocation process and 5) providing follow-up instruction to collection services. Since this parallel run will begin in June 2011 a summary of our findings will be shared with the work session participants.

## **VI. References**

Ravindra, Daniela (2009). *Integrated Business Statistics - Program Blueprint*. Statistics Canada. Internal document.

Bricault, Mario (2010). *Common Editing Strategy – Discussion Paper*. Statistics Canada. Internal document.

Bricault, Mario Saint-Pierre, Étienne (2011). *The common editing strategy and the data processing of business statistics surveys - Invited Paper*, United Nations Statistical Commission and Economic Commission for Europe - Conference of European Statisticians, Work Session on Statistical Data Editing, Ljubljana, Slovenia, 9-11 May 2011

Bocage, Yannick Godbout, Serge (2011). *Quality and Efficiency Using a Top-Down Approach in The Canadian Integrated Business Statistics Program – Macro Editing Methods- Supporting paper*, United Nations Statistical Commission and Economic Commission for Europe - Conference of European Statisticians, Work Session on Statistical Data Editing, Ljubljana, Slovenia, 9-11 May 2011

Brodeur Marie, Doherty, Karen Dufour Stéphane, Lussier Robert, Mayda Jackey, McAuley Julie, Ménard Mario, Ridgeway Art, Royce Don, Smith Wayne (April 2009). *Report of the Task Force on Corporate Business Architecture*. Statistics Canada, Internal document

Brodeur Marie (2006) et al, *The integrated approach to economic surveys in Canada*, Statistics Canada, Catalogue 68-514.