

**UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE**

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Budapest, Hungary, 14-16 September 2015)

Topic (ii): Managing and supporting changes related to editing and imputation.

Redefining roles and responsibilities in a new harmonized statistical production process: opportunities and challenges

Prepared by Etienne Saint-Pierre, Statistics Canada

I. Introduction

1. Profound changes to statistical processes, organizational structures and systems infrastructure were successfully implemented at Statistics Canada during the past five years. These changes ensured that relevant data of high quality could continue to be produced during a time of financial constraints. One of the largest changes implemented was a complete overhaul of the methods and operations used to produce economic data at Statistics Canada. The new economic statistics model is called the Integrated Business Statistics Program (IBSP). Currently, standardized processes, methods and tools have been applied to 70 IBSP programs and an additional 80 will be incorporated into the model by 2019.
2. Statistics Canada is moving away from a traditional model where all the statistical production processes for a given program are under the responsibility of the same organizational unit. This shift started fifteen years ago with the Unified Enterprises Statistics Program (UES)¹ but accelerated in the last five years. In the IBSP, ten or so different internal service centers are involved in the various statistical production processes.
3. At the same time, the IBSP led to the introduction of innovative and harmonized statistical processes that changed previously used practices. Hundreds of employees have had to adjust the way they work after their programs were integrated into IBSP.
4. In Section II, the highlights of the new operational framework used for business surveys at Statistics Canada will be described.
5. These changes have led to a complete reassessment and readjustment of the roles and responsibilities of the actors of the statistical processes: from survey specific specialist to process specific specialist; from a silo type management approach to a purely matrix environment. Using the Generic Statistical Business Process Model, the evolution of roles and responsibilities for methodologists, analysts and technical support staff over the last few years will be described in Section III. The acceptance and

¹ For more details on the UES, see Brodeur et al. (2006)

understanding of these new roles by all concerned employees and the efficient management of the many and complex interrelations between all partners are among the most important post-implementation challenges in IBSP.

6. Challenges and opportunities of the new business model will be discussed in the section IV and V of the paper using concrete examples from two key processes of the new operating model: the design of electronic questionnaires and the edit and imputation process.

II. New operational framework for Business Surveys

7. In 2010, Statistics Canada launched the Corporate Business Architecture (CBA) initiative. At the time, there were growing financial pressures in the organization which led to a thorough review of business methods, statistical processes, and systems infrastructure. The main CBA objectives are to harvest efficiency with minimal cuts to statistical and analytical programs, determine methods for enhancing the quality of statistics including enhancing quality assurance, and finding ways to improve responsiveness in the delivery of statistical programs.

8. The Report of the Task Force on Corporate Business Architecture (2009) also recommended the mandatory use of these Corporate Services for all statistical programs and projects:

- Methodology service
- Business and institutional frame service
- Questionnaire design resource centre
- Collection service
- Operations service
- Data service centre
- Internet dissemination service
- Inquiry service
- Processing service
- Informatics Service
- Administrative Data Division

9. It was recommended to also create additional generic corporate services when the pooling of skills could create economies of scale.

10. The development and the implementation of the new operational framework for processing business surveys was driven by key CBA principles: Corporately optimal decisions supersede locally optimal decisions; develop metadata driven processes; reuse business processing and analytic systems across programs to minimize the number of processing systems that need to be maintained.

11. The IBSP provides a common framework for the various business programs conducted at Statistics Canada. By 2019, nearly 150 programs (annual, sub-annual, industry specific, economy-wide, activity based) will be integrated into this new harmonized framework. The key elements of the new model are:

- Mandatory use of Statistics Canada's Business Register as a common frame;
- Electronic questionnaires as the principal mode of collection;
- Production of estimates in an iterative fashion and active assessment of the quality of the data to improve collection management;

- Mandatory use of generalized systems for sampling, imputation, estimation, application of confidentiality and dissemination;
- Maximum use of tax and auxiliary data to replace collected data from respondents;
- Implementing an Integrated Infrastructure System based on a flexible metadata driven model;

12. Integration to the IBSP led to new roles and responsibilities for most subject-matter and service provider divisions. For programs previously integrated to the UES, the changes were relatively minor. Getting their managers to embrace the IBSP corporate services oriented approach proved to be relatively easy. For other programs, the changes in roles and responsibilities were far more complex. These changes and their associated challenges have been a challenge for some managers.

III. Changing roles and responsibilities in the statistical business process model

13. The use of generic corporate services throughout the statistical process has led to a complete change of roles and responsibilities. These changes are illustrated in Figures 1 and 2. For each sub-process of the Generic Business Statistical Process Model (GSBPM), the service center that is primarily responsible is identified (however several service centers are involved in each process). As shown in Figure 1, the coordination of the statistical processes involved eight different service centers under the IBSP in comparison to four or five for the majority of surveys outside the IBSP (Figure 2). Few other existing surveys have currently only one or two service centers involved in the entire statistical process.

Figure 1- Service Centre which holds the primary responsibility **under the IBSP** by sub-process of the GSBPM model

Quality management							
Metadata management							
Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
Identify needs	Design outputs	Build collection instrument	Create frame and select sample	Integrate data	Prepare draft output	Update output systems	Gather evaluation inputs
Consult and confirm needs	Design variable descriptions	Build or enhance process components	Set up collection	Classify and code	Validate output	Produce dissemination products	Conduct evaluation
Establish objectives output	Design collection	Build or enhance dissemination components	Run collection	Review and validate	Interpret and explain outputs	Manage release of dissemination products	Agree on action plans
Identify concepts	Design sampling frame and sample	Configure workflows	Finalise collection	Edit & Impute	Apply disclosure control	Promote dissemination products	
Check data availability	Design Processing and Analysis	Test production system		Derive new variable & units	Finalise outputs	Manage user support	
Prepare business case	Design production systems and workflow	Test statistical business process		Calculate weights			
		Finalise production systems		Calculate aggregates			
				Finalise data files			
Center of expertise							Shared responsibilities
Subject-Matter Division	Metadata Service	Content Specifications Service	Methodology Service	Processing Service	Collection Services	IT Service	Dissemination Service

Figure 2 - Service Centre which holds the primary responsibility by sub-process
 (Representative of most business surveys **outside the IBSP**)

Quality management							
Metadata management							
Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
Identify needs	Design outputs	Build collection instrument	Create frame and select sample	Integrate data	Prepare draft output	Update output systems	Gather evaluation inputs
Consult and confirm needs	Design variable descriptions	Build or enhance process components	Set up collection	Classify and code	Validate output	Produce dissemination products	Conduct evaluation
Establish objectives output	Design collection	Build or enhance dissemination components	Run collection	Review and validate	Interpret and explain outputs	Manage release of dissemination products	Agree on action plans
Identify concepts	Design sampling frame and sample	Configure workflows	Finalise collection	Edit & impute	Apply disclosure control	Promote dissemination products	
Check data availability	Design processing and analysis	Test production system		Derive new variable & units	Finalise outputs	Manage user support	
Prepare business case	Design production systems and workflow	Test statistical business process		Calculate weights			
		Finalise production systems		Calculate aggregates			
				Finalise data files			
Center of expertise							
Subject-Matter Division	Methodology Service	Collection Services	IT Service	Dissemination Service			

14. Under the IBSP, subject-matter divisions’ **primary** responsibilities are concentrated at the beginning (specify needs) and at the end (analysis and dissemination) of the statistical business process. While they are consulted and encouraged to actively participate in specifying requirements and solutions for other processes, they are no longer responsible for delivery. Their role in the Build, Collect and Process components of the GSBPM is significantly reduced as Statistics Canada moved towards a specialization of responsibilities through centers of expertise, each doing their share in each program.

15. The subject-matter divisions are the first point of contact with data users and they have to ensure that statistical needs are addressed through existing or new programs. They are also responsible for data analysis and correcting anomalies in the data. They now spend more time on analysis and outreach to improve the relevance of data outputs, while relying on centres of expertise to collect and process data.

16. With regards to the Methodology Service, there is a shift in responsibility from the Process to the Design component. Methodology resources are systematically involved in both the methodology and processing framework Design for all business surveys. The process functions in which they were involved are transferred to a centralized Processing Service. In such an integrated framework and with so many players involved, this is needed to ensure the optimality and consistency in all processes and between surveys. Quality measures must be defined and integrated into every step of the statistical business

process. The Methodology Service plays a leading role in managing the overarching quality assurance for the different processes.

17. The creation of a specialized team for the management of IBSP metadata, the mandatory shift for all data processing activities from subject-matter divisions or Methodology Service to the Processing Service and the centralization of all IT activities are also among the main changes of the new business model.

18. The management of the statistical process now requires a coordinated delivery by all partners in the matrix environment. The transition to this new business model is a challenge for the survey managers who used to interact with a limited number of service centers. The decision making process now involves a large number of stakeholders. Governance structures, management committees and a fully integrated schedule are in place to ensure that local preferences are not detrimental to corporate goals in all decisions.

19. The various challenges and opportunities of using multiple corporate services in the statistical business process will be discussed with concrete examples for two key components of the IBSP: the design of electronic questionnaires and the edit and imputation (E&I) process. The role of each service center will be described.

IV. Challenges and opportunities - Design of Electronic Questionnaires

20. Designing and building an electronic questionnaire in the IBSP requires the involvement of several centers of expertise.

21. Once the content of a survey is confirmed by the **Subject-Matter division**, after consultation with users or the survey sponsor, a questionnaire is developed with the input of the **Methodology Service** to ensure that the questionnaire design and content is optimal for the edit and imputation (E&I) process. The preliminary design is reviewed closely by the **Questionnaire Design Resource Center (QDRC)**, which ensures that collection vehicles are standardized and that the questionnaire length is reasonable for the respondents. This Center then goes on the field to test the questionnaire and assess respondent's ability to answer the questions. Then, the **Metadata Service** generates formal cell numbers, mnemonics and variable names in accordance with the concepts measured and the actual electronic questionnaire is developed by the **Electronic Questionnaires Development Service**.

A. Opportunities

- (a) The review of questionnaire designs by the Methodology Service led to efficiency gains. The questionnaire content and design is optimized to reduce response burden and ensure the E&I system will have the best possible inputs for dealing with partial non response. The addition of totals and subtotals to the electronic questionnaire, the creation of derived variables, the elimination of non-essential cells, the grouping of variables and the design of better grids are examples of recommendations that help to facilitate processing. The resulting design ensures data can be processed in the generalized imputation system with streamlined metadata, minimal transformations, and optimized input for dealing with partial non response.
- (b) All questionnaires must meet high quality standards for their questions, guides, accessibility and propensity to measure the desired concepts. The experience of the respondent is also at the heart of the questionnaire design. QDRC and the Electronic Questionnaires Development Service have

developed an expertise that can be applied to many surveys.

- (c) The coherence between surveys is favored as the Metadata Service ensures that similar variable names and cell numbers are assigned for similar concepts across all surveys in the Business Statistics Program. This facilitates the use of cross-survey data as an auxiliary source for E&I as well as for coherence analysis.
- (d) The development of electronic questionnaire specifications requires navigating through several templates and the use a very specific syntax. It would be very expensive to have each survey specialist learn to develop specifications and to perform this task. The centralization of this task has led to efficiency gains, enhanced the quality of the product and reduced the number of players required to interact with the Electronic Questionnaires Development Service.

B. Challenges

- (a) *Balance corporate goals and local needs:* To keep their programs relevant, the Subject-matter and National accounts divisions revise survey content regularly. This entails changes to questionnaires, addition of new sources of administrative data or changes in the processing and dissemination of the information to satisfy user needs. In the context of a highly centralized production model, these needs have to be assessed against those of all other stakeholders’.

It is therefore no longer possible for an area to unilaterally decide to implement change. For all proposed changes, a Change Management Committee, with representatives from each service center assess the relevance, impact, cost and timing of the suggested changes based on existing priorities and capacities. This group ensures that decisions are not driven solely by individual survey preferences, but are justified and aligned with corporate goals.

Finding the right balance and reconciling competing needs represents a challenge. A continuous prioritization process has to happen to ensure that, overall, the organisational priorities are met. This is managed through a well structured governance process lead by a Project Management Team made up of senior executives from all stakeholder areas. It is responsible for strategic decisions and high level issue resolution. A number of interconnected committees, responsible for specific areas of operation, report to the Project Management Team

- (b) *Selecting the optimal option when the costs are not equally absorbed by all service centers:* There is often more than one option available to address non standard requirements; selecting the most effective option at the corporate level requires a thorough review of the pros and cons. Such a review is challenging as stakeholders typically negotiate to maximize local efficiency, considering their own limitations and requesting services be delivered by other stakeholders. Seeking consensus for the most effective corporate solution requires strong governance.

An IBSP example of such a scenario occurred when it became necessary to agree on a solution for converting reported data to a common unit of measure for editing, imputation and estimation. For Subject-Matter Divisions, responsible for the management of response burden, the prescription of a common unit of measure on the questionnaire is not an acceptable solution as it might have a negative impact on response propensity. From that point of view, offering as many options as possible is the optimal scenario. The Collection Services team argued that conversion of data should be done in post collection processes. From their point of view, implementing conversions in the collection system would require the development of survey specific functionality to transform raw data, which is not their mandate. The perspective of the Processing Service experts was that not converting units of measure in the collection system would be

detrimental to data quality as online historical collection edits could not be appropriately triggered when respondents changed reporting unit of measure from one cycle to the next. In summary: strong but conflicting arguments could be made depending on the various points of view.

In most cases, the chosen option will involve additional costs in terms of time and effort to one of the service centers. The negotiations and the cost-benefit analysis of the various options are thereby more complex since they involve multiple players incurring different costs. Roles and responsibilities are often part of these discussions. Who is ultimately responsible for the conversion of data into standard units? Respondents, Collection Service or the Processing Service? Compromises and the strong IBSP governance are key to ensure that decisions are optimal according to corporate objectives.

V. Challenges and opportunities - Define and apply edit and imputation strategies

22. Every survey that is part of the IBSP must have automated edit and imputation strategies designed by the Methodology Service, no matter the size of the survey or the number of respondents. The use of the generalized imputation system is also mandatory. It will no longer be possible to have an in-house solution for the imputation process. Three service centers are directly involved in the E&I process.

23. First, the **Methodology Service** and the **Subject-Matter Division** discuss the outline of the E&I strategies based on the content of the questionnaire, the subject matter of the survey and the data sources available. Among other things, they review the relationship between variables, identify requirements for deriving variables to meet specific needs of data users, and review existing strategies for resolving failed edits by referencing the best auxiliary inputs available.

24. The **Methodology Service** then defines a detailed imputation strategy that makes optimal use of all the existing functionalities and methods of the generalized imputation system as they are the ones with in-depth knowledge of this system. They translate the E&I strategy into metadata templates specifying sequences of tasks, edit groups, imputation classes, rules for deterministic imputation, hierarchy of methods, etc . They also define strategies to ensure automated resolution of all failed edits before estimates are produced.

25. After the detailed strategies and templates are completed, the **Processing Service** enters all the metadata into the processing application that contains validation rules. Metadata tables for E&I have numerous and complex interdependencies with many other metadata tables related to the definition of variables, the integration of data and the estimation process. They must be entered with a very specific syntax, according to a defined sequence. Coherence must be preserved across the fifty or so metadata tables. The strategies are then tested by the **Processing Service** and the results are validated by the **Methodology Service** before they are implemented in production. Adjustments to strategies can be easily put in place through the various iteration of testing or production iterations.

26. Once implemented in the production environment, the **Processing Service** processes data in an iterative and fully automated mode, which is referred to as Rolling Estimates². In the rolling estimates process, data move from integration to E&I, to the production of estimates and quality indicators. The output of each iteration is used to automatically readjust follow-up priorities by **Collection Services**. Results from the various iterations are analysed by the Subject Matter divisions to identify anomalies and

² The Rolling Estimates data processing model is defined in Godbout, Beaucage, Turmelle (2011)

correct irregularities. This review often leads to adjustments to the E&I strategies. The **Processing Service** can easily modify the parameters when changes are recommended by the **Methodology Service**.

A. **Opportunities**

- (a) This model ensures that all IBSP Surveys have imputation strategies defined by the Methodology Service anchored on the best use of the generalized imputation system and within the constraints imposed by the harmonized framework.
- (b) By design, this model is not conducive to manual interventions. It is no longer required to do manual editing of data before processing. Automated solutions are defined to replace manual editing strategies previously put in place by Subject-Matter Division. In the Rolling Estimates model, estimates and quality indicators are generated before the editing tool can be accessed. The interface supports micro/macro adjustments following top down analysis. Equivalencies, sum of parts, linear relationship between variables, zero filling, balancing, and comparison with predicted values are all aspects of the system. Most of the manual interventions previously done before imputation have been automated.
- (c) This model does not require an extensive knowledge of the E & I methodology and the supporting metadata for the Subject-Matter Division. Their involvement in this process is to highlight the key elements that may influence the strategies and to correct anomalies that may arise. This model does not require any special knowledge of the Subject-Matter for the Processing Service. The work is the same regardless of the survey. Their expertise is transferable from one survey to another. This is a source of efficiency. Currently the ratio in the Processing Service is equal to 3 surveys / person and is expected to be around 5 surveys / person by 2019. A total of 10 surveys a day on average will be processed by the Processing Service in 2019.
- (d) In several programs, the Methodology Service was not only responsible for defining the imputation strategies but was also involved in the transformation of the collected data prior to their integration in the imputation systems. In these programs, they were also responsible for building, running and changing programs for the imputation component. In the new model, all production functions are transferred to the Processing Service. The role of methodologists is now to provide methodological support, optimize the quality assurance process and measures, to do research or to develop models. The Methodology Service is no longer doing or supporting production.
- (e) The fully metadata driven process reduces the dependency on IT and on specific knowledge of individuals. Adjustments can be made to imputation strategies simply by having the Processing Service modify the metadata.

B. **Challenges**

- (a) *Sign-off on strategies designed / built externally:* Subject-Matter Divisions are asked to sign-off on imputation strategies defined by the Methodology Service and to certify the final data. It is difficult for Subject-Matter Division employees to approve a strategy defined in dozens of metadata tables with technical vocabulary in a very specialized syntax. It is also difficult for them to verify that all quality control mechanisms were put in place and properly enforced in the many steps of the E&I process as this process is not run by them. It is also difficult for the employees to know if the many metadata entries reflect the defined strategy. Some survey managers feel as if they need to know every detail to validate and approve strategies as they are ultimately responsible for the disseminated data. It can lead to a lot of time and effort spent understanding

all the details of E & I process.

To overcome these challenges, accountability must be shared between the different service centers, each one being accountable for its own area of responsibility. In coming years, quality control processes, performance metrics and service level agreements will be developed to overcome some of these challenges. The notion of sign-off is no longer as binding as it used to be as the strategies can be modified very quickly by simply adjusting the metadata between each iteration.

- (b) *Different interpretation of the concept of quality:* According to Statistics Canada Corporate Business Plan: “*The pursuit of quality is not a question of achieving arbitrarily high standards, whatever the cost, but rather of ensuring that the standard achieved is appropriate given the intended uses of the information*”³. The Methodology Service has the mandate to define E&I and estimation strategies designed to produce estimates of quality that is “Fit for purpose”.

The Methodology Service’s interpretation of the “*Fit for purpose*” concept might not be shared by the Processing or the Subject-Matter Divisions thus bringing up debates about the level of precision and complexity of the methods and tools to build - complexity that often come with a high marginal cost. There are working groups and steering committees in place to guide and intervene when the various centers are not in agreement on the notion of “Fit for purpose”. Again the strength of the governance structure is key for optimal decisions.

The IBSP framework has been designed to maximize the precision and minimize the bias of the estimates at the macro level. The tools, by design, have been built to support a top-down analysis approach and not to perform efficiently a high volume of manual interventions at micro level. For some analysts, the notion of “Fit for Purpose” quality is tied up to a bottom-up approach where the quality of estimates is a direct function of the number of verifications / interventions at micro-level. Typically this approach requires the production of multiple reports, the inclusion of numerous flags or the addition of functionalities to facilitate the detection and correction of errors and to identify potential incoherencies after their manual interventions at micro level. They would also like to consider the inclusion of a pre-grooming stage before the E&I process – which would mean the design of a costly alternative flow to the existing model.

It is important to increase the level of ‘trust’ that analysts have in the process ---letting the system and strategies defined by the Methodology Service do the heavy lifting. Training and consultations are essential to make sure analysts understand that most of the pre-grooming they currently do can be automated. The E&I process also preserve the coherence inside a record.

- (c) *Define the responsibilities at the intersection points:* It was challenging to define clearly responsibilities for certain tasks at intersection points of sub-processes. Stakeholders tended to define their responsibilities by specifying the functions that were not part of their responsibilities. For example, The Collection Services considered they were not responsible to build zero-filling functionalities or a unit conversion module as they are not responsible for transforming raw data. The Processing Service considered they are not responsible to map and transform auxiliary files to complement E&I strategies, consequently, they argued that Subject-Matter Divisions should do it. This led to some confusion and debate on the role and responsibilities for completing specific tasks. As the project moves forward, consensus is established and new business rules are defined.

³ Statistics Canada Corporate Business Plan 2014-2015 to 2015-2016

- (d) *Determining priorities when the demand exceeds the capacities:* In order to minimize the cost of respondent follow-up activities, the Collection Service needs to frequently have an updated list of priorities in order to reach the quality targets with a minimum number of calls. New reporting units and analyst's interventions in the data can modify these priorities. Subject-matter Divisions frequently ask for runs of rolling estimates to see the impact of both their interventions and the new data available on the estimates. As more and more data are available, the Methodology Service will sometimes recommend adjustments to the E&I strategies and will request additional runs of rolling estimates to review the results.

The burden of producing the rolling estimates, troubleshooting and modifying processing parameters and testing of strategies is the responsibility of the Processing Service. When demand for Processing services exceed capacity, priorities must be set. What criteria should be used to determine that one survey is more urgent than the other? Who ultimately decides on the priorities?

Survey managers that are new to the integrated framework often express concerns with the timeliness of producing estimates. By outsourcing data processing, survey managers fear the loss of responsiveness.

To overcome this challenge, iterations of rolling estimates are planned in advance to ensure the system has the capacity to deliver. Stakeholders participate in operations management committees where decisions are made to address conflicting priorities. The IT Service also works to increase system capacity.

VI. CONCLUSION

27. The new harmonized processing framework introduced with the IBSP is anchored on the use of multiple Corporate Services for the various statistical processes. This led to a complete overhaul of the roles and responsibilities of the various actors in the statistical process. While the benefits of this new processing model and revised roles and responsibilities are clearly seen in the 70 integrated surveys, there are still numerous challenges that could weaken the support of survey managers, analysts and corporate partners to the IBSP. This support is absolutely essential to the success of the large-scale implementation of the new harmonized model for business surveys.

28. The large number of actors involved directly or indirectly in the survey process under the IBSP makes the coordination of all activities and the decisions making process far more complex than when all the business processes were managed more locally.

29. A strong governance enforcing the promotion of corporate goals over local preferences, well-defined service level agreements, performance and quality metrics for the various service centers are key to success. Once a few cycles of the new programs are successfully completed, confidence in the new model will increase and roles and responsibilities will become clearer.

VII. References

Brodeur M., Koumanakos, P., Leduc, J., Rancourt, E., Wilson, K. (2006). *The integrated approach to economic surveys in Canada*. Statistics Canada Catalogue no. 68-514. Ottawa

Godbout, Serge, Yanick Beaucage and Claude Turmelle (2011). *Quality and Efficiency using a Top-Down Approach in the Canadian's Integrated Business Statistics Program*. Conference of European Statisticians, Work Session on Statistical Data Editing. Ljubljana, Slovenia (9-11 May 2011).

Statistics Canada. *Corporate Business Plan, 2014-2015 to 2016-2017*. www.statcan.gc.ca

Statistics Canada. (2015). *Integrated Business Statistics Program Overview*. Statistics Canada Catalogue no. 68-515-X. Ottawa

Statistics Canada. *Report of the Task Force on Corporate Business Architecture* (2009). Internal Document.