The Integrated Business Statistics Program Is Ten Years Old!
A Methodological Perspective

Pierre Daoust¹, Matei Mireuta¹, Gildas Kleim¹

ABSTRACT

Statistics Canada introduced in 2010 the Integrated Business Statistics Program (IBSP) with the aim of achieving greater efficiency in processing Statistics Canada’s business surveys, while maintaining or improving the quality of data. Since its deployment in 2013, the IBSP has provided a standardized framework for over a hundred and twenty economic surveys. IBSP surveys use Statistics Canada’s Business Register as a common frame and collect data primarily through electronic questionnaires that are based on harmonized concepts and content. Surveys share common sampling, collection and processing methodologies that are driven by metadata and Statistics Canada’s suite of generalized systems. In addition, common tools are in place to edit, correct, and analyze data. This article provides a methodological perspective on the ten years of experience at Statistics Canada of integrating and processing economic surveys under this standardized framework.

KEY WORDS: Economic surveys, Integrated business statistics program, Standardized survey framework, Statistics Canada

1. INTRODUCTION

As a result of the Corporate Business Architecture (CBA) review initiative that was launched in 2009, Statistics Canada started a general integration of its business surveys. First deployed in 2013 on around 40 surveys, the Integrated Business Statistics Program (Statistics Canada, 2015) now provides a standardized framework for 128 economic surveys on topics as diverse as agriculture, manufacturing industries or services.

The development of IBSP required a wide range of new approaches, including methods to take into account complex unit structures and multiple allocation objectives at the sample design stage, orienting the active collection efforts to optimize the quality of key estimates, measuring the variability due to imputation for the survey estimates, and expanding the usage of calibrating estimates on known totals extracted from secondary data sources. The methodology for IBSP has been presented in several articles, the reader can consult (Godbout, 2011), (Godbout et al., 2011), (Turmelle et al., 2012), (Mills et al., 2013), (Turmelle and Beaucage, 2013), (Turmelle et al., 2014), (Andrews et al., 2016), (Mireuta et al., 2017), (Andrews et al., 2018), and (Daoust et al., 2018). In the present article the authors will discuss how Statistics Canada adjusted the IBSP methodology over the years considering the various challenges encountered, the evolving needs of surveys, the developments in statistical norms and generalized systems, and to the emergence of modernization initiatives. Finally, the article highlights future strategic orientations and challenges.

The second section will summarize the IBSP’s objectives, as well as the principles that led to its implementation. The third section will draw a profile of the surveys that are currently integrated in IBSP, while the fourth will describe the strategic methodological principles. The fifth section will expose the main challenges IBSP had to face overtime, and the solutions that were found. Future directions will be formulated in the sixth section.

¹ ESMD, 100 Tunney’s Pasture Driveway, Ottawa, Ontario, K1A 0T6, pierre.daoust@statcan.gc.ca, matei.mireuta@statcan.gc.ca, gildas.kleim@statcan.gc.ca.
2. OBJECTIVES AND MAIN PRINCIPLES

The development and architecture of IBSP were based on the following 6 core objectives and guiding principles.

2.1 Improve Data Quality

Improvement of data quality was to be achieved by implementing standardized processing methods and tools, as well as harmonized questionnaire content (for example, by asking the same set of questions on expenses across all the different business surveys), and by facilitating coherence analysis.

2.2 Reduce Response Burden

Providing accurate answers to the questions asked during a survey is a long and fastidious process for businesses that can result in an important economic cost, and Statistics Canada has always aimed at reducing it. The development of IBSP gave the opportunity to adopt more efficient techniques for that purpose. This led to the use of more efficient sampling techniques and strategies, the extensive use of administrative data, or by setting collection strategies and tools aiming at limiting collection operations to essential units, as well as stopping them when specific objectives have been met.

2.3 Modernize data processing infrastructure

The Unified Enterprise Survey (also known as the UES) upon which the IBSP was founded had aging infrastructure and was not compatible with electronic questionnaires (EQ) and many of the Statistics Canada’s generalized systems. There were also numerous workarounds and manual interventions that were required for processing. The development of IBSP gave the opportunity to modernize this infrastructure.

2.4 Integrate the majority of Economic surveys

The UES included 60 annual surveys that were integrated regarding their content, collection activities and data processing. In the late 2000s, a need emerged at Statistics Canada to increase this number and develop a system that would include a greater diversity of surveys, such as cost-recoverable surveys and sub-annual programs.

The IBSP integration took place in several waves, going from around 40 surveys at wave one in 2013 to nearly 130 surveys in 2023. For each survey, the strategies for all survey steps are specified in standardized metadata files and tested before going into production.

2.5 Standardize processes and tools

At the time the CBA was conducted, UES was at a point where survey-specific needs were met at the expense of global ones: although most surveys used similar approaches and methodologies, these steps were conducted independently across the surveys, with no general oversight or communication between respective survey teams. As a result, it had become very difficult to bring modifications to the system for fear of destabilizing it. The IBSP was to address this issue by using or developing a set of standard methods and software that could be used across all surveys.

2.6 Flexibility to adapt to change

While using standard techniques and tools was one of IBSP’s goal, it was acknowledged that a certain level of flexibility needed to be kept. The main challenge for the IBSP was to come up with a survey design strategy and infrastructure that:

- Could integrate the majority of economic surveys, as well as potential ad hoc surveys;
- Focuses on improving the quality of commodities and characteristics estimates, given the extensive use of tax sources for financial data;
- Offers the flexibility to add extra sample, new questions or produce new estimates, as needed;
- Easily adapts to new auxiliary data, classifications, etc.

The IBSP processing system itself makes use of modular approaches at each step of the survey, giving the user the ability to easily assess different solutions. Another desirable feature of the system is the ability to replace modules by new ones. This way, new methods can be introduced without changing the entire system, only adapting affected modules and their related metadata.
In time, additional features or methodological approaches have been added to the IBSP system. This is often the number one requirement from the survey areas. As this will be discussed in the fifth section, the IBSP system was able to adapt to specific situations. In some cases, surveys adapted to IBSP standards, while in others, the IBSP adapted to new realities.

3. PROFILE OF SURVEYS INTEGRATED IN IBSP

3.1 Subject Matter Area and Size

The surveys now integrated in IBSP span a variety of diverse subject matter areas as well as a variety of sample sizes (Figure 3.1). Surveys having mainly financial variables (Retail, Services, Manufacturing, and Investment) were among the first surveys to join the IBSP program over a decade ago. However, over the years, surveys with different information needs such as energy output, farming output, public transport ridership and so on have been successfully integrated and illustrate the versatility of the IBSP system. Through an elaborate metadata management system, surveys in the same subject matter area share common variable names, definitions, and concepts. The IBSP system is now also quite versatile in terms of the size of the processing files that can be successfully managed, ranging from several units to millions of units.

![Figure 3.1: Subject Matter Areas and Size Profiles of Surveys Integrated in IBSP](image)

3.2 Processing Frequency and Sampling Design

Almost 90 annual surveys are processed in IBSP and this represents the most common processing frequency. In addition, almost 25 surveys are processed in IBSP on a monthly basis. The remainder of surveys integrated in IBSP have other sub-annual processing frequencies (quarterly, triennial, semi-annual) or processing frequencies exceeding one year (quinquennial).

The surveys now integrated in IBSP include 67 censuses and 43 classical sample surveys. Since reference year 2018, 18 surveys using a hybrid sampling strategy have also been processed in IBSP (these are referred to as pseudo-censuses). This hybrid strategy is essentially an administrative census that is complemented by survey data for a sample of the population.
4. IBSP STRATEGIC METHODOLOGICAL PRINCIPLES

The IBSP methodology is based on strategic principles that are aligned with the six IBSP core objectives described in section 2. Several of these principles will be described below.

4.1 Standardization of Methodology and Systems

4.1.1 Metadata Driven Model

In the IBSP system, metadata are stored in easily modifiable tables that are used to drive systems programs. An important advantage is that changes required as program needs evolve can be accommodated by modifying metadata, rather than by rewriting system code. This provides more control for the processing team and more flexibility for users.

4.1.2 Common Statistical Standards and Tools

The IBSP system is built on the concept of a harmonized content model. Therefore, surveys must apply statistical standard classifications such as the North American Industrial Classification System (NAICS)\(^2\) to classify the target population by industry, the North American Product Classification System (NAPCS)\(^3\) to categorize and collect business input and output data, and many others. In addition, many concepts and variables are shared across IBSP surveys and these have been harmonized to create common definitions and to implement common survey content.

Furthermore, processing in IBSP is predominantly carried out using Statistics Canada’s suite of general systems such as G-SAM for sampling (Statistics Canada, 2013), BANFF for edit and imputation (Statistics Canada, 2017), and G-EST for estimation (Statistics Canada, 2019).

4.2 Electronic Questionnaires and Administrative Data

Several initiatives have been implemented in IBSP in order to improve collection efficiency. For example, the electronic questionnaire serves as a common and main collection instrument for IBSP surveys. Furthermore, the use of administrative data to replace collection data is fully integrated in the IBSP system. For the vast majority of surveys, tax data are used as replacement for financial variables for a subset of the survey population comprised mainly of medium and small enterprises.

4.3 Rolling Estimates Model

The methodology of IBSP is adapted to the Rolling Estimates Processing model which is automated, iterative and runs regularly during the cycle of a survey until an acceptable level of quality is achieved (Godbout et al. (2011)) (Figure 4.1). Briefly, the first step of a rolling estimate iteration consists in importing collection data, administrative data and auxiliary data. Then, all required processing steps are performed to create preliminary estimates. At this stage, the quality of a few important (or key) estimates is assessed against predetermined quality targets. Several quality indicators are available in IBSP, but surveys predominantly use a quality indicator based on a weighted response rate coupled with a quality indicator based on deviations from predicted values. At the same time, an impact score is calculated (measure of impact) for each unit and represents the anticipated increase in quality that would be observed if the unit was successfully followed up. These impact scores drive subsequent active collect management as well as active analysis management activities and have an impact on the data available to the next rolling estimate iteration.

---

\(^2\) North American Industry Classification System (NAICS) Canada 2022 Version 1.0 (statcan.gc.ca)

\(^3\) North American Product Classification System (NAPCS) Canada 2022 Version 1.0 (statcan.gc.ca)
4.4 Calibration and Estimation of Total Variance

One important goal of IBSP was to expand the use of calibration as part of the estimation strategy of integrating surveys. As a result, the use of calibration (generally to tax revenue) has significantly increased for economic surveys. Another important goal of IBSP was to implement a methodology to estimate the total variance of estimates, i.e., the variance due to sampling and the variance due to imputation. The variance due to imputation is estimated with Statistics Canada’s System for Estimation of Variance due to Nonresponse and Imputation (SEVANI) (Beaumont and Bissonnette (2011)) and quality indicators based on total variance are now produced for several economic surveys.

4.5 Two-Phase Sampling

The development of IBSP saw the implementation of a novel two-phase sampling approach for annual economic surveys. The main goal of the first phase was to refine industry classification of units, in-scope/out-of-scope status for units in a few activity-based surveys and to update commodity data for trade surveys. The two-phase sampling strategy saw significant practical challenges and was eventually abandoned as discussed in the next section.

5. KEY CHALLENGES FOR IBSP METHODOLOGY

A key to success in developing and implementing IBSP was the ability to identify and solve key challenges. This section will focus on some of those challenges related to IBSP Methodology. Several of these were identified in advance, and mitigation strategies were put in place, others emerged over the years and IBSP needed to adapt to the situation. These are internal challenges in Statistics Canada, associated to the tasks of adapting most economic surveys to a new standardized framework or adapting this framework when required.

5.1 General challenges

First some general challenges for IBSP that also impacted the methodology are presented, and the corresponding strategy adopted to overcome them.
5.1.1 Challenges

Resource capacity:

- The integration into IBSP requires a significant amount of survey teams’ time (analysts and methodologists) from the consultation stage through to the survey integration stage, including development of specifications/metadata, input of files and testing and review of outputs. This often interfered with production priorities which almost always took precedence.
- In addition, it was often the case that survey teams were integrating numerous surveys at the same time putting a further strain on resources.

Timeline to define processing strategies:

- Many of the surveys processing specifications were developed many years ago and were sometimes not well documented, not fully understood by the current survey teams, not aligned with IBSP methodology, or improvements were required to reflect the programs’ needs. In those cases, it implied that survey teams needed to start from scratch when defining processing strategies, although this also offered opportunities to improve the methodology.

Timeline between consultation and production:

- With a 2-to-3-year timeline from consultation to production, many of the original staff involved in the consultations left, leading to new requests or requirements. These were not only a result of staff rotation, but also due to new demands from survey managers to align to evolving programs’ needs. The response to these requests was particularly complicated by the fact they often were made once surveys had started integration or once key integration decisions had already been taken.
- When consultations began, the IBSP system was still in major development for the first waves of surveys integrating IBSP. This presented the challenge of collecting requirements and defining solutions without a stable infrastructure in place. As time passed, the consultation process was easier as the infrastructure itself was better defined and more mature.

Global perspective versus survey specific:

- Changing from a more survey specific perspective where survey teams had their own custom strategies and methodologies, tailored to manage very survey specific needs was not straightforward as the teams felt they were losing flexibility and local optimality.

Resistance to change:

- Sometimes survey managers had been working with their surveys for numerous years and were reluctant to change. This made gathering requirements and meeting timelines difficult. In that context there were a few roadblocks to overcome:
  o Strong belief in the uniqueness and non-standard nature of the program,
  o Fears that survey teams’ needs will not be met,
  o Uncertainty as to the new roles and responsibilities.

Schedules flexibility:

- Delays to one survey meant working on multiple surveys at the same time. Schedules were initially created to optimize managing multiple surveys at the same time. With numerous surveys to integrate at the same time changes to timelines created significant bottlenecks for key resources.

5.1.2 Mitigation strategies

Strong governance and support from management:

- Each functional level had a higher level of recourse going from the working level to the project management team consisting of directors and director generals, where key issues impacting several surveys would be resolved.
Integrating into the IBSP was strongly supported by management. Switching to a more generic approach meant change and sometimes new or different approaches to work. Management would also need to be involved in convincing staff that this approach was better for the corporation, even if it meant local sacrifices.

Dedicated project managers:

- In certain cases, the parent divisions of the survey teams would have a dedicated person responsible for providing the information needed to the transition. This helped get more compliance as survey teams were more encouraged to work with someone from their own division. This also reduced some of the survey teams’ workload.

Training:

- Training was a key piece of the process that started during the consultation and continued during the transition. It was available in both formal and informal settings. It was very important in ensuring that people understand the concepts, the integration process, and the governance structure to make the transition easier.

Strong and constant communication / Managing timelines and schedule monitoring:

- Regular transition meetings took place with all partners to ensure staff were aware of upcoming tasks, deadlines, roles and responsibilities, and highlight/explain task details. This meeting was also used to monitor and update the schedule when required.
- Updates to the schedule were sent to all partners and their management to ensure all levels were aware of any delays or issues.

Knowledgeable transition staff:

- One of the key pieces was to have dedicated IBSP experts that understood all aspects and could guide everyone throughout the integration process. These IBSP experts were available to assist everyone to figure out the best approaches and solutions to problems.

5.2 Surveys requesting unique strategies

5.2.1 Challenge

Surveys aimed first at adapting to IBSP standards, but in some cases asked for unique strategies not yet supported by IBSP. These could occur to replicate pre-IBSP strategies, to respond to emerging needs for their program, or in line with modernizing initiatives for their program.

5.2.2 Mitigation strategy

IBSP used a business case approach in these cases, that led to 4 different outcomes:

- The survey found a way to adapt to IBSP standards
- Creative solutions were suggested with minimal impact on IBSP. For example, the standard survey IBSP methodology is being used to implement pseudo-censuses mentioned in section 3.2.
- New methods were developed or added in IBSP. An example of this is surveys wishing to take advantage of new functionalities added to the generalized systems.
- In some cases, part of the processes was done outside of IBSP. For example, a few surveys with unique special sampling methodology needs have their sample selected outside of IBSP but then integrated in IBSP for collection and processing.

5.3 Creating/improving the sampling/processing strategies

5.3.1 Challenge

Trying to facilitate the development of the sampling and processing strategies, to avoid costly adjustments during production, was a key aspect to consider for IBSP. Conducting studies to define these strategies outside of IBSP, and then integrating those in IBSP, had the risk of being time consuming and inefficient.
5.3.2 Mitigation strategy

Several improvements were made to address this:

- **IBSP Sandbox environments** were created to develop the sampling, active collection and estimation strategies (and such an environment is currently being developed for E&I). This allowed survey teams to compare different strategies using directly the IBSP systems and fine-tune their methodology before implementing it in production.
- **Common tools** were developed to support studies. This includes detailed diagnostics to assist in validating results, dashboards (for active collection) to help visualize the relevant information to users, and extension of the sandboxes to test potential improvements to the IBSP methodology.
- **Advanced templates** were added to support Edit & Imputation (E&I). The E&I metadata content was extensive, and initially converting the E&I strategy into metadata was very time consuming. The advanced templates, designed in Excel, greatly improved the users’ experience and reduced the risk of errors.

5.4 Two-phase approach for annual economic surveys

5.4.1 Challenge

The two-phase sampling approach for annual economic surveys, described in section 4.5, was a major initiative for IBSP. Its efficiency was thoroughly evaluated, and a few challenges were identified:

- The first phase information on commodities was analyzed for 3 years of processing (2013 to 2015). There were relatively high nonresponse rates of the first phase data on commodities, and insufficient treatment of the responses (for coding of write-ins) that could be completed ahead of the second phase, and therefore the benefits were found to be relatively small.
- A detailed analysis of the impact of the first phase between 2013 and 2018 showed that the small benefits obtained by the two-phase design could not compensate for the challenges that came with the much more complex environment this created and the cost for this approach.
  - The Statistics Canada business register information on industries used by industry-based surveys was confirmed to be of very high quality based on the first phase responses, and the first phase was subject to relatively high nonresponse rates. This resulted in negligible benefits for these surveys except for a few industries.
  - The two-phase design was useful for two activity-based surveys, leading to some gains on the precision or coverage of these surveys. However, the corresponding survey teams indicated that these surveys could produce reliable estimates with a few enhancements in a one-phase design context.
  - The two-phase approach complicated the design, implementation, and validation of surveys. Analysts were often challenged to determine if the variation between cycle in the estimates was due to fluctuations related to this design or were real data changes.
  - It delayed the start of collection for the detailed questionnaires of the second phase and impacted the timeliness of releases.
  - It sometimes led to survey teams inflating the number of units selected with certainty (leading to some loss of efficiency).

5.4.2 Mitigation strategy

The challenges were thoroughly discussed using IBSP governance, and the following decisions were made:

- After the 2015 processing, IBSP temporarily stopped using the module on commodities in the first phase electronic questionnaire, with an aim of retooling for a future restart.
- Studies were conducted to determine the details of the methodology required to use a one-phase strategy and predict the impact for key variables.
  - These concluded that the impacted surveys were ready to move to a one-phase design, noting that it would improve the precision of estimates, reduce the incidence of extreme weights for some surveys, and improve the stability of estimates between years.
  - Therefore, all the annual economic surveys reverted to one phase designs after 2018 processing.
There were some enhancements made to the design of these surveys at that time (for example changes were made to the sampling design to improve the stability between cycles, and an enhanced usage of secondary data was introduced).

5.5 Completing the development of advanced functionalities of QIMI

5.5.1 Challenge

Several factors made it very challenging to develop advanced functionalities for the Quality Indicators and Measures of Impact System. The two major improvements are model driven, for which it is difficult to respect some of IBSP principles, such as offering standardized processes and tools (and avoid manual interventions as much as possible). These are:

- Adding the total CV (including variance due to imputation) as a quality measure. The objective is to align the quality measures used in QIMI with the key quality indicators for IBSP. Some of the challenges in achieving this are adapting the methods to ensure the reliability and stability of the model-driven approach during collection, determining how this quality measure could be used in combination with other quality measures, and working collaboratively with survey teams to help them understand and be comfortable with this approach.
- Adding to QIMI the response propensity dimension. The objective would be to enhance QIMI by estimating during collection the probability that the unit will respond, whether there is a follow-up attempt or not. Again, the challenges are to develop a methodology that will be reliable throughout the collection period, that wouldn’t rely on manual interventions, and that can be well understood by survey teams.

5.5.2 Mitigation strategy

Study and research are being conducted to add these dimensions. These are being integrated in a major initiative at Statistics Canada in the context of improving collection methodologies for business surveys.

5.6 Calibration

5.6.1 Challenge

Adding calibration was new to most of the business surveys. Survey teams asked for some assistance to decide when it was beneficial to add this step to their estimation strategy.

5.6.2 Mitigation strategy

A few actions were taken to assist surveys in deciding on calibration.

- A working group of experienced methodologists made recommendations to assist in making these decisions.
- Calibration results were thoroughly analyzed in the first few years of IBSP and worldwide experts from the Advisory Committee of Statistical Methods at Statistics Canada were consulted to make improvements to the methodology.

5.7 Total variance

5.7.1 Challenge

Given the innovative nature of this for most surveys, it has been very challenging for them to start using quality indicators based on total variance (that includes variance due to imputation). Also, such quality indicators are currently not available for all variables. Several surveys are conducting small studies initially to find out the impact this will have on their quality indicators, and to suggest modifications to their E&I strategies.

5.7.2 Mitigation strategy

The following actions were needed.

- New tools were added to support impact studies aimed at preparing surveys on this
- New developments are required in SEVANI and IBSP to extend the methodology to all variables.
- Strong support was obtained through the IBSP governance structure, to reinforce the priority of adopting quality indicators that relies on total variance.
6. FUTURE PRIORITIES FOR IBSP METHODOLOGY

IBSP has successfully adapted over the years to respond to the surveys' needs and will continue to focus on known priority challenges. However, the next few years will be very challenging in relation to two major initiatives at Statistics Canada.

First, all sectors at Statistics Canada will be asked to focus on the conversion of the existing programs to an open-source coding environment. This impacts IBSP from different angles. On one hand, it will be a major undertaking as IBSP is an extensive and complex system and has several interdependencies with other important systems at Statistics Canada. On the other hand, it will open the door to opportunities, to adapt quickly to advanced methodologies and to enhance partnership prospects. For more information on the conversion of Statistics Canada programs to open-source coding please see Burnett-Isaacs (2023).

Second, there is an ongoing key agency review at Statistics Canada aimed at modernizing our statistical programs so that they remain efficient and connected with the ever-evolving needs of our data users and partners. In doing so:

- IBSP will need to continue to provide users with information on the quality of the data using modern methods.
- There might be a need to assess whether Statistics Canada should continue to adapt IBSP to evolving modern methods and tools or whether it would be better to develop and build a new environment. There might be a need to re-evaluate the guiding objectives and principles of IBSP to adapt to the requirements of modern programs.

ACKNOWLEDGEMENTS

We would like to thank Claude Turmelle, Alexander Reicker, and Beatrice Baribeau for reviewing this paper and for their helpful comments.

REFERENCES


