# Current and Future Applications of the Generic Statistical Business Process Model at Statistic Canada

**Q2010 Conference, Helsinki, Finland, May 2010**

Laurie Reedman and Claude Julien
Statistics Canada
150 Parkdale Avenue
Ottawa, Canada
Laurie.Reedman@statcan.gc.ca Claude.Julien@statcan.gc.ca

Keywords: data quality, efficiency, communication

## 1 Introduction

The Generic Statistical Business Process Model (GSBPM) was first developed by Statistics New Zealand, and was revised at the Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS) in 2007, 2008 and 2009. The intended use of the GSBPM is to provide a basis for statistical organizations to agree on standard terminology to aid their discussions on developing metadata systems and processes.

Statistics Canada is a centralized agency. Its head office houses all program management and processing functions, as well as methods and systems research and development. This centralization allows the use of a tool such as the GSBPM to further integrate, standardize and streamline processes. This paper describes how the GSBPM is used in recent initiatives to ensure the data quality of its products and the efficiency of its statistical programs.

On ensuring data quality, Section 2 describes how the model has been used as a framework for reviewing several statistical programs for their quality assurance practices and identifying which sub-processes have a greater risk of errors in quality. Section 3 illustrates how the model has proven to be effective in supporting the analysis, summary and communication of corrections made to *The Daily*, the Agency's official release bulletin. Section 4 describes how the model has facilitated communication between the managers responsible for producing Statistics Canada's fifth edition of the *Quality Guidelines* and the managers of statistical programs by providing a comprehensive graphical representation of the basic statistical process.

On the efficiency front, Statistics Canada is currently facing the challenge of maintaining the quality of its information products while using fewer resources. In response to this challenge, a Corporate Business Architecture initiative has been launched. As part of this iniative, the GSBPM has been integrated into an information management model for statistical information holdings, as described in Section 5. It is within this context that Statistics Canada is looking at significantly optimizing and standardizing the processes involved in the production of statistical output. The proposal to develop an Integrated Business Statistics Program (IBSP) represents the way to achieve these objectives in business statistics. As described in Section 6, the recommendations for the IBSP

embrace the use of the GSBPM to align services, align underlying information technology/information management with business strategies and describe business processes using common terminology.

## 2 Quality Assurance Reviews

At Statistics Canada, every statistical program, as well as all infrastructure and operations programs, must carry out an extensive self-assessment on its financial performance (efficiency), data quality (relevance, accuracy, interpretability, timeliness, coherence, and accessibility), management of human resources and strategic outlook every four years to produce a Quadrennial Program Report. Program areas also produce a follow-up biennial report. Since 2007, Statistics Canada has also developed and implemented an annual quality review program during which quality assurance practices in a set of statistical programs are subjected to an independent internal assessment. All of these reports and assessments are presented to the Policy Committee (Statistics Canada's topmost management committee).

In the first two waves of reviews, the focus was on assessing the measures taken in the execution of programs to prevent erroneous data from being released to the public, i.e., to assure the accuracy of the data. The objectives of the review were to identify best practices that should be promoted to other programs at Statistics Canada as well as areas where action is needed.

Nine statistical programs were reviewed in 2007 and six more in 2008. A small team of reviewers was assigned to each program to be reviewed. Each team was provided with all relevant existing documentation about the statistical program, and also met with program managers and team members to discuss the program's quality assurance practices. A detailed questionnaire was developed to guide the discussions, but was provided only as an aid and was not intended to be restrictive. It covered all the steps in a typical statistical program and posed questions designed to stimulate discussion of the strengths and vulnerabilities of the program with respect to the accuracy of its data products. The risks identified in the first two waves of reviews fell into several broad categories: staffing, project management, systems and processing, documentation and dissemination. Some of these categories cut across all stages of a program (staffing, project management, documentation) while others referred to specific stages (processing, dissemination).

In 2009, a third wave of reviews was undertaken. This time the scope of the review was broadened to include risks in the design of the program as well as risks in its automated systems. The discussions included not only program managers and team members but also service and internal data providers, such as methodologists, system developers and representatives from areas responsible for administrative data such as the Business Register Division and Tax Data Division.

To guide the discussions, a tool was needed that would be easier to use than the long questionnaire. This tool would have to ensure that no steps in the program would be overlooked. Also, it was deemed useful to provide the reviewers with knowledge of

some of the higher risk areas identified in the first two waves. The GSBPM fit these needs admirably. The programs reviewed in 2009 included large- and small-scale survey programs, as well as programs dependent on administrative data. The GSBPM is designed to be independent of the data source, and it can therefore be used to describe processes based on surveys, censuses, administrative records and other non-statistical or mixed sources. When this model was presented to the reviewers, the sub-processes where risks had been identified in previous reviews were highlighted, indicating that while all steps should be covered in the review, certain ones might stand out. Below is the model as it was provided to the reviewers.

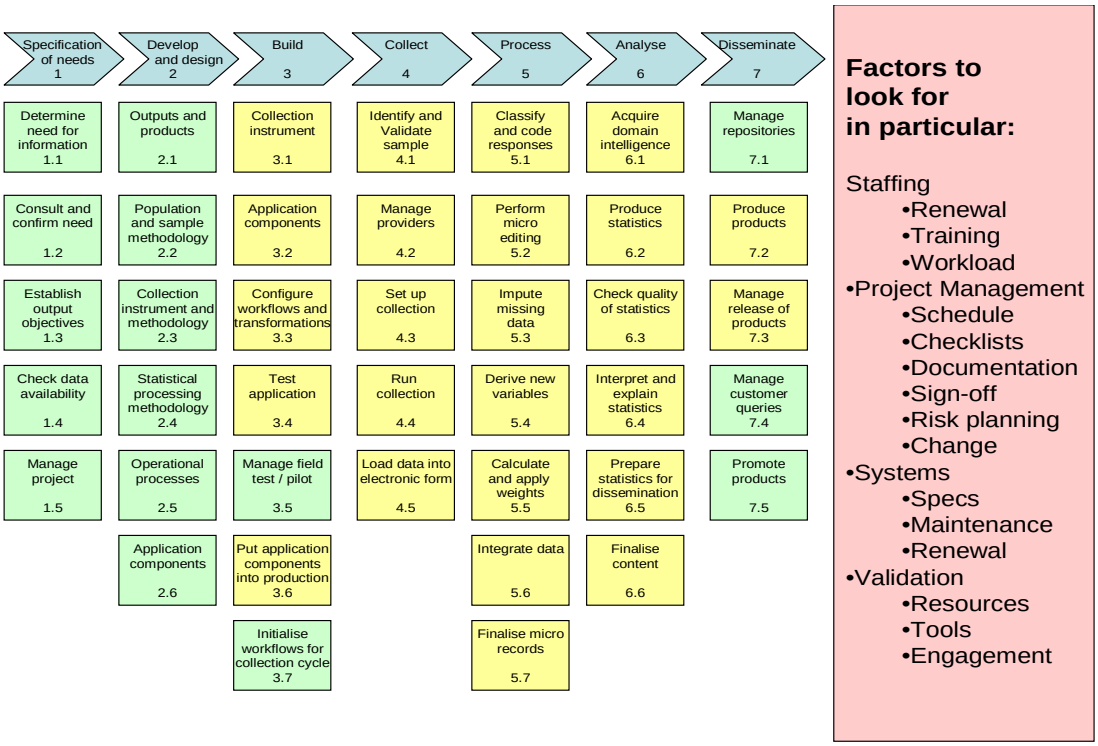


Figure 1: Model used in Quality Assurance Reviews

The GSBPM was shared with program managers and team members when they met with the reviewers to discuss their program. In most instances, a printed copy of the GSBPM was on the table as an "aide memoire" throughout the meeting. The model facilitated communication by providing a comprehensive graphical representation of the basic statistical process.

Each review resulted in a report detailing the risks along with mitigating actions that were framed as best practices if already in place and working well, or as recommendations if not. By following the logical order of steps in a statistical program as set out in the GSBPM, the reader could immediately orient the risks and compare the location of risks of one program to another. Using the common language as prescribed by the model also facilitated communication with a broader audience.

The quality reviews as well as other reporting mechanisms surface many issues that would appear to be a high priority at the local level. Putting these issues into the context of the GSBPM enables us to identify the global issues for which a solution would have broad applicability. Used in this way, the GSBPM is a tool for improving the efficiency and effectiveness of quality improvement initiatives.

## 3 Analysis of corrections

Statistics Canada conducts nearly 400 statistical programs. *The Daily* is Statistics Canada's official release bulletin and first line of communication with the media and the public. *The Daily* issues news releases on current social and economic conditions and announces new products. It provides a comprehensive one-stop overview of new information available from Statistics Canada.

Most of the data releases are accompanied by an article that varies from a brief announcement that data are available to an in-depth presentation and analysis of the data. Over 1,250 articles are published every year in *The Daily*, or 5 per day on average. Producing these articles is a fairly complex process. Many articles contain tables and charts. Most articles contain links to previous releases, metadata or more detailed information that are released at the same time (e.g., analytical papers, time series tables, etc.). All articles include a contact for obtaining more information. Furthermore, the articles and all related publications, time series tables and metadata are released at the same time. The production and concurrent dissemination of all this information requires the effort of many employees in different areas of the organization working under tight deadlines. This effort includes a series of checks conducted by the communications staff prior to release.

Given the amount of information released and the complexity of the release process, and in spite of the numerous checks carried out, minor corrections are occasionally made to the articles after their release. In June 2007, Statistics Canada set out to record every correction made to articles dating back to July 2006 and thereafter. Several basic characteristics are recorded:

- Date of release;
- Date of correction (to analyze the delay in correcting the articles);
- Statistical program (e.g., Price index, Manufacturing, Health);
- Frequency of release (e.g., monthly, annual, analytical study);
- Location of correction in the article: text, tables, chart or contact information;
- What was corrected: words, numbers, reference period, link or metadata.

With this basic information, the corrections are divided into two groups: those that affect the accuracy of the data (e.g., corrections to numbers in a table) and those that do not. Examples of corrections that do not affect accuracy are corrections to the name or telephone number of a contact person and incorrect links to metadata or more detailed information. Corrections for accuracy are further investigated to gather the following information:

- The magnitude of the correction;
- How the error was detected or reported;

- Where in the process the error occurred;
- Which underlying risk factors may have led to the error.

The GSBPM provides a good map on which to pin the origin of errors and identify the "hot spots" or steps of the process that are more at risk. Like the quality reviews described in Section 2, the recording and analysis of corrections for accuracy started before the GSBPM made its way into our operations. Notwithstanding, the origin of the corrections recorded so far can be easily located in one of the nine phases of the statistical business process (Level 1).

**Table 1 – Corrections for accuracy**

| Year | Texts | Correction Rate | Location of error and relative magitude of correction | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Design | | Build | | Collect | | Process | | Disseminate | | Don't know | |
| | | | Higher | Lower | Higher | Lower | Higher | Lower | Higher | Lower | Higher | Lower | Higher | Lower |
| 2006 | 623 | 2,1% | 1 | | | | | 1 | | 4 | 1 | 3 | | 3 |
| 2007 | 1234 | 3,1% | | | | | 3 | | 6 | 14 | 2 | 10 | | 3 |
| 2008 | 1260 | 2,2% | | | 1 | | 1 | 2 | 2 | 10 | 1 | 11 | | |
| 2009 | 1171 | 0,8% | | | | | 2 | | | 1 | 2 | 1 | 3 | |
| 2010 | 315 | 0,3% | | | | | | | | | | | | 1 |

The analysis of corrections for accuracy based on the Level 1 phases of the GSBPM has highlighted the following:

- It is clear that errors originating at the later phases, especially the dissemination phase, have a smaller magnitude (e.g., corrections in the order of one or two tenths of a percentage points in the text or in a table). These errors were more likely to occur when we started recording the corrections. This analysis raised awareness throughout the Agency and managers took action to improve quality assurance measures and overall communication to address the issue.
- On the other hand, errors originating at the earlier phases (e.g., collection) are much less likely to occur but have more impact.
- As a result of this analysis and other measures taken, the percentage of texts with a correction for accuracy has decreased significantly since 2007.
- The results of the analysis are coherent with the outcome and evolution of the quality reviews. The errors that were occurring at the later phases have decreased. The current errors have been more concentrated at the earlier phases, as early as the design and build stage. Accordingly, the focus of the quality reviews has also broadened and shifted to the earlier phases.

In the context of this operation, the GSBPM provides a comprehensive map and common terminology needed to efficiently record and summarize a set of fairly rare events. This allows us to pinpoint riskier phases or sub-processes and communicate them to program and senior managers. As we gather evidence from both the quality reviews on programs and errors leading to corrections, we can overlay their respective statistical process risk maps to efficiently identify, prioritize and address quality issues.

## 4 Quality Guidelines, 5th Edition

Statistics Canada's management structure, policies and guidelines, consultative mechanisms, project development and management approach, and environment have been developed to facilitate and assure effective management of quality. The basic

mechanisms for managing quality are described in Statistics Canada's Quality Assurance Framework. The framework consists of a wide variety of mechanisms and processes acting at various levels throughout Statistics Canada's programs and across its organization. The effectiveness of this framework depends not on any one mechanism or process but on the collective effect of many interdependent measures. These build on the staff's professional interests and motivation. They reinforce each other as means to serve client needs. They emphasize Statistics Canada's objective professionalism and reflect a concern for data quality. An important feature of this strategy is the synergy resulting from the many players in Statistics Canada's programs operating within a framework of coherent processes and consistent messages. Within the framework, the Quality Guidelines (Statistics Canada, 2009) provide an accompanying document that describes a set of best practices for all of the steps of a statistical program, and is aimed at project team members who are charged with the development and implementation of statistical programs.

Underlying all of these mechanisms, processes and practices are eight guiding principles. One of these principles is the following: quality must be built in at each phase of the process.

Statistical agencies have often modeled the statistical process, as an aid in managing it. The second edition (April 1987) of Statistics Canada's Quality Guidelines contained a schematic of the statistical survey process. More recently, the fifth edition (October 2009) contains the 2008 draft version of the GSBPM.

These various models have in common the division of the process into a number of phases or steps. While the details of the various models vary, all contain common elements: the specification of user needs, the design of the program, the implementation or "build" phase (systems, operations manuals, training), the execution phase (collection, verification) and the evaluation phase. A basic principle of quality assurance is that it must be considered at all of these phases. If the user's needs are not understood or are incorrectly specified, then all the steps that follow will only result in data that are not fit for the user's intended use. Measures must also be taken to ensure that the design is done properly; if it is not, then no amount of perfect implementation and execution will compensate. However, having a good design is not enough; if it is not implemented or executed correctly, then the good design has gone to waste. Also, without proper evaluation, the statistical agency will not know whether or not the statistical program has met its objectives.

The principle that quality must be built into each phase, together with the notion that quality is multidimensional, leads logically to the conceptualization of quality assurance management as a matrix defined by the dimensions of quality (relevance, accuracy) in one dimension and the phases of a survey (specification of needs, design, implementation, execution, evaluation) in the other dimension. Presented below is the GSBPM where mauve shading has been used to indicate the Level 2 steps for which there are recommendations in the Quality Guidelines 5$^{th}$ Edition. Given that the Quality Guidelines have been written primarily by and for methodologists, it is not a surprise that

there are more guidelines for the early steps. Future editions of the Quality Guidelines should be expanded to cover more of the Level 2 steps. To improve the accessibility of the guidelines themselves, future editions could be structured around the model, and the reader could locate the corresponding guidelines by navigating through the model itself.

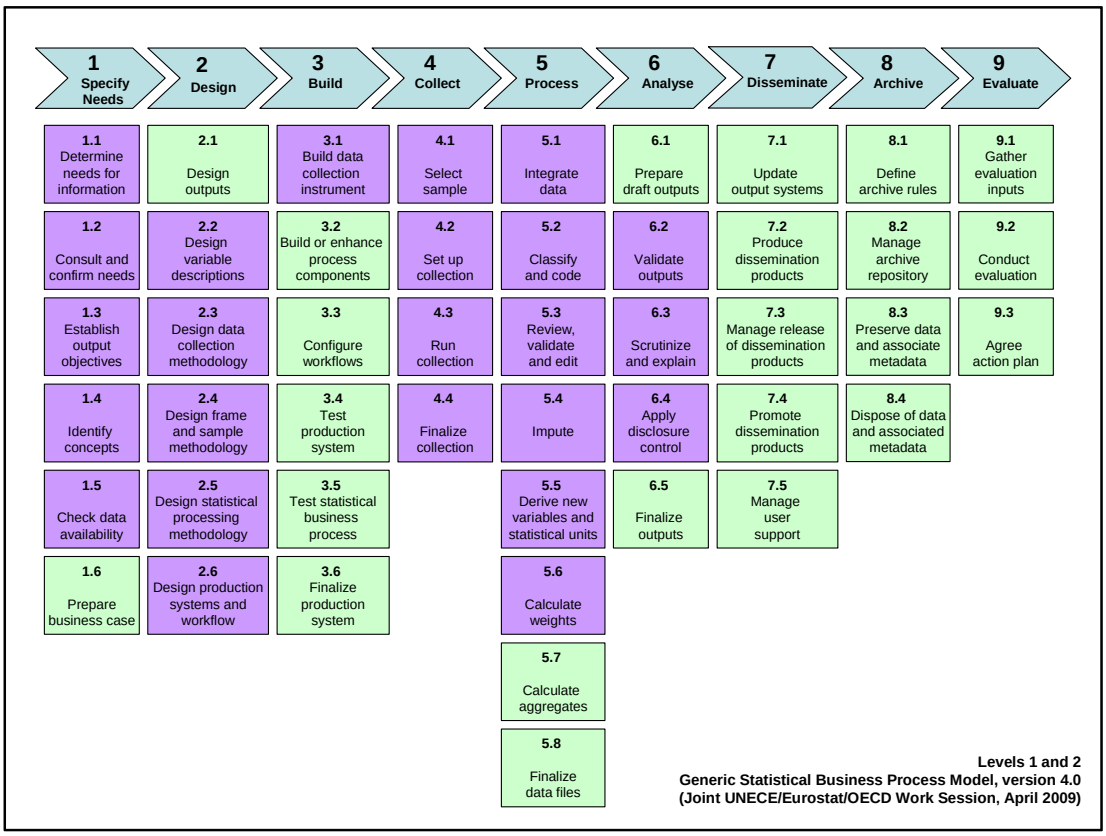

Figure 2 – Steps in GSBPM for which there are quality guidelines

## 5 Corporate Business Architecture

Statistics Canada faces numerous challenges in maintaining the quality of its information. The economy and society we are attempting to measure is changing at an unprecedented pace. Household and business respondents are less willing to participate in surveys, while their lifestyles and technology make it increasingly difficult to contact them. Critical tools and computer systems must be maintained, and, where it is important to our clients, international standards must be met and international comparability of information maintained.

Statistics Canada has also been asked to deliver its program with significantly fewer resources. Combined with the challenges of improving relevance while maintaining quality, this requirement will force Statistics Canada to achieve a much greater efficiency than at any time past.

The Corporate Business Architecture responds in part to the challenges facing Statistics Canada by addressing significant aspects of the quality and cost dimensions. The

initiative will conduct a comprehensive review and revision of Statistics Canada's business architecture in order to achieve or advance three objectives:

- A harvestable efficiency on ongoing operating costs to meet corporate downsizing commitments with minimal cuts to statistical and analytical programs;
- Enhance quality assurance through the implementation of a reduced, unduplicated set of more robust systems and processes that are properly maintained and documented; and
- Improved responsiveness in the delivery of new statistical programs by streamlining Statistics Canada's core business process.

Pending further action on the proposal, a taskforce was created to consider what measures could be taken in the short term in order to advance the objectives of the Corporate Business Architecture initiative while generating efficiencies in the short to medium term. To structure its analysis and organize its report, the taskforce has drawn on the GSBPM. A version of this business process model is embedded in the following diagram under the heading Core Business Process.

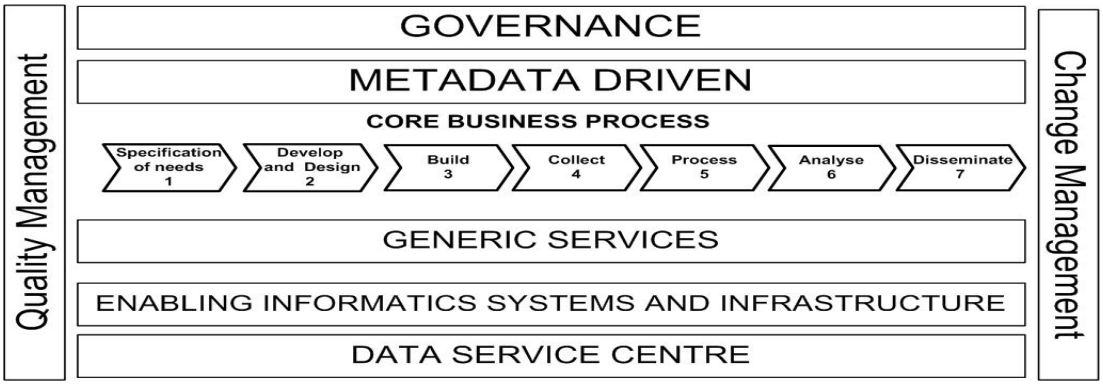

Figure 3 – Model of Corporate Business Architecture

The report from the taskforce goes on to describe each box in the above model and make recommendations. The report explains the concept of the data service centre identified in the illustration and referred to repeatedly in the recommendations. The data service centre, whether there be one, two or more, uses a common set of technologies and tools to store, organize and provide access to statistical information to support processing, dissemination and analysis. The data service centre develops and applies an information management model for statistical information holdings. The data service centre can thus represent any statistical program at Statistics Canada, and the core business process phases likewise represent those in any statistical program.

## 6 Integrated Business Statistics Program

In the spring of 2009, Statistics Canada started planning a new approach to producing business statistics at Statistics Canada. This new approach, named the Integrated Business Statistics Program (IBSP) is to be based on the existing model of the Unified Enterprise Statistics program (UES) which had been conceptualized in 1996 and implemented through different phases up until 2008.

The UES marked an important new development for Statistics Canada in its business survey programs. Rather than meeting demands for increasing the quality and quantity of available business statistics through a scaling-up of its existing surveys, Statistics Canada opted for a general restructuring of its business statistics programs. This restructuring aimed to bring greater integration, consistency and coverage while improving operating costs and allowing for better management of response burden.

Currently, the UES consists of 60 surveys which are integrated in terms of content, collection and data processing. The surveys were integrated slowly over a period of approximately 12 years, and the model has worked fairly successfully. However, at this juncture, due to a desire to increase efficiency, robustness and responsiveness to outside demand, Statistics Canada would like to take the UES to the next level by developing a model that would apply to a much larger array of business surveys than the UES currently does.

A software engineering and architecture team was created and tasked with designing a business solution that would align with the corporate business survey processing vision. Among its deliverables was a set of key recommendations to address the ongoing issues faced in the UES and contribute to the long-term objectives set out for the IBSP.

The first recommendation was that the IBSP formally identify and define the necessary services to support statistical business processing from the survey planning to dissemination phases. More specifically, it was recommended to align IBSP services with the GSBPM as prescribed by the Corporate Business Architecture. The model provides a set of standard services to support processing of business statistics. Among the benefits of using a service model would be a common understanding of the "big picture", better management of overall survey processing, the use of human resources across divisions and the optimization of overall business statistical processing processes.

The second recommendation was that IBSP develop and maintain a formal Business Process Model (BPM) by providing a step-by-step description of the interaction between business processes and organization units toward the accomplishment of a single objective. A BPM is an extension of the service model where each service is opened up to discover the business scenarios and actors. The BPM is also a very useful tool to align underlying information technology/information management with business strategies.

The third recommendation was to make use of corporate services, statistical processing standards, industry best practices and the Corporate Business Architecture principles wherever possible. This would include having the IBSP business processes described using the GSBPM, which is based on recognized statistical processing standards for business statistics. An obvious benefit of this would be an improvement in communication through the use of a standard common language.

## 7 Conclusions

The Generic Statistical Business Process Model (GSBPM) has proven to be a very useful addition to Statistics Canada's quality assurance practices. The broad applicability of the

GSBPM has made it a very useful tool for focussing the analysis of risks to quality and communicating where stronger quality assurance practices are recommended. The model will very likely support future initiatives and play an important role in planning and structuring the next edition of the Quality Guidelines. It has also become a central tool in developing statistical metadata systems and processes. It will provide the basic process structure and terminology with which Statistics Canada will be conducting a comprehensive review and revision of its business architecture. Finally, Statistics Canada is planning to expand the quality reviews on statistical programs to a broader review on the quality and efficiency of statistical processes used across many programs. The model will be crucial in ensuring that such reviews are conducted in a well-coordinated and efficient manner.

**REFERENCES**

Government of Canada, Department Performance Report – Statistics Canada, Treasury Board Secretariat, http://www.tbs-sct.gc.ca/dpr-rmr/2008-2009/index-eng.asp?acr=1529

Statistics Canada (2009) "Quality Guidelines (Fifth Edition)", October, 2009.

UNECE Secretariat (2009) "Generic Statistical Business Process Model: Version 4.0 – April 2009", Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS).