11-617B
no. 2008-003
c.1

Statistics Canada Statistique Canada

NOT FOR LOAN
NE S'EMPRUNTE PAS

Methodology Branch

Business Survey Methods Division

Direction de la méthodologie

Division des méthodes d'enquêtes-auprès des entreprises

Canada

WORKING PAPER NO. BSMD-2008-003E/F
METHODOLOGY BRANCH

CAHIER DE TRAVAIL NO. DMEE-2008-003F/E
DIRECTION DE LA MÉTHODOLOGIE

Quality Review of Key Indicators at Statistics Canada / Examen de la qualité des indicateurs à Statistique Canada

Claude Julien, Don Royce

BSMD-2008-003 E/F

STATISTICS CANADA STATISTIQUE CANADA
APR 14 2008
LIBRARY
BIBLIOTHÈQUE

Quality Review of Key Indicators at Statistics Canada

Claude Julien, Don Royce

Abstract

In September 2006, Statistics Canada formed a committee to conduct a review of the quality assurance practices in the production of nine key indicators (e.g. employment, consumer prices, gross domestic product, etc.). The review was carried out on a variety of business, economic and household surveys based on respondent data, administrative data or other survey outputs (i.e. derived statistics). The review was completed in March 2007. This paper describes the review process and summarizes its key findings. It also highlights the benefits of the review to the statistical programs, the individuals involved in the review and Statistics Canada's Quality Assurance Framework.

Keywords: Quality Review, Quality Assurance

1. Introduction

Statistics Canada has a long history of sound quality management practices, including numerous policies, standards and guidelines that define how it communicates with users, plans its resources, designs and conducts its surveys, disseminates its data and evaluates its activities. In preparation for an audit on quality by the Office of the Auditor-General (OAG) in 1997, Statistics Canada brought together its quality management practices into a Quality Assurance Framework (QAF). The QAF (Statistics Canada, 2002 and Brackstone, 1999) defines data quality in terms of six dimensions (relevance, accuracy, timeliness, coherence, interpretability and accessibility) and describes how each dimension is managed.

The framework was used to carry out a self-assessment of four programs for the purpose of the audit on quality. The OAG was very satisfied with the self-assessments and recommended that they be applied more widely. Statistics Canada has since put in place a formal integrated program reporting mechanism (IPR) under which each program reports on quality, costs, management of human resources and strategic directions. Each program carries out an extensive review every four year (quadrennial report) and a shorter follow-up review in between (biennial report). The review is mostly in the form of a self-assessment, but a user satisfaction survey or similar type of consultation is also required.

The IPR focuses on the assessment of a statistical program's outputs. In the last year or so, Statistics Canada has been conducting research into adding another review process to its reporting system. This review process would focus on the program processes rather than program outputs. Review strategies in other organizations, such as Statistics Sweden's Quality Audits (Eiderbrandt-Nilsson, 2004), U.S. Census Bureau audits (Bushery, McGovern, Marck and Hogan, 2006) and the Development of a Self-assessment Program (DESAP, Laiho and Nimmergut, 2004) have been considered. Julien and Born (2006) presented this research at European Conference on Quality is Survey Statistics - Q2006.

This research was abruptly accelerated when, in September 2006, in response to a small number of unfortunate errors in data released in the preceding eighteen months, Statistics Canada decided to undertake a review of the quality assurance practices in its statistical programs. Section 2 of this paper discusses some of the challenges in designing the review and the key decisions that had to be made. Section 3 describes how the review was implemented. Section 4 highlights its main findings. Section 5 continues with the lessons learned on the review process. Section 6 concludes with future research on developing an on-going review process that could be adapted to different programs and different needs of the organization.

2. Design of the Review

Statistics Canada's Policy Committee, composed of the Chief Statistician and the Assistant-Chief Statisticians, identified nine programs to be reviewed. There were six monthly programs: Labour Force Survey (LFS), Monthly Retail Trade Survey (MRTS), Monthly Survey of Manufacturing (MSM), Consumer Price Index (CPI), International Trade (IT) and Gross Domestic Product by Industry (GDP). There were three quarterly programs: Income and Expenditure Accounts (IEA), Labour Productivity (LP) and Balance of Payments (BOP). These programs were selected, not because it was believed that they were particularly prone to errors, but because they represent the most "mission critical" sub-annual programs of Statistics Canada, providing current information on economic conditions. The quality of these "key indicators" is therefore of utmost importance to a wide range of users.

To oversee the review, a Steering Committee was formed consisting of the four Directors General from Methodology, Economy-wide Statistics, System of the National Accounts, and Labour and Household Surveys. The Steering Committee was chaired by the Director General of Methodology who called upon the Quality Secretariat, also located in Methodology, to provide assistance in developing and coordinating the review.

With direction from Policy Committee, the Steering Committee established its objectives and developed an approach to review these programs within a very tight time schedule. The specific objectives of the review were twofold:

i. to identify any specific weaknesses where action is needed, and the underlying factors behind them, and
ii. to identify "best practices" that should be promoted to other programs.

This balanced approach, looking at both positive and negative aspects, was a key principle of the review. The identification of the underlying factors was crucial to allow senior management to optimally allocate resources to address the quality issues raised by the review.

The review was to be completed by the end of the calendar year, i.e. within four months. Given this very short timeframe, it was quickly decided to conduct the reviews simultaneously. This led to two considerable design challenges: developing a single review process that could be quickly and effectively applied to each program; and clearly defining the scope of the review, i.e. the aspects of the program and the dimensions of quality to review.

The first design challenge came from the considerable variety in the characteristics of the statistical programs. LFS, MSM and MRTS are "traditional" surveys with a selection of sampling units and collection of data from respondents; yet the former is a household survey and the other two are business surveys. The CPI is a less traditional survey compared to the three others with collection of information by field staff and from other sources. IT is a survey based uniquely on an extremely high volume of administrative data. The GDP, IEA and LP are derived statistics programs. They do not conduct surveys on their own, but collect statistical information from other programs that they then transform and analyze to produce a more coherent and comprehensive set of statistics. Finally, BOP incorporates characteristics of all these types of programs. This variety made it challenging to develop a common process to review each program.

The scope of a quality review is nearly unlimited. Given the short timeframe for the review, it was essential to lay out all the aspects of the programs and quality that could be reviewed. Firstly, it was clear from the initial mandate from Policy Committee that the review was to focus on the program processes rather then their outputs. As mentioned earlier, the program outputs are already assessed by the integrated program reports. Secondly, a statistical program process is comprised of a series of steps (e.g. setting objectives, defining a population, creating a frame, selecting a sample, etc.). Thirdly, different dimensions of a particular step can be reviewed: its design (how the step is laid out on paper), its implementation (how the design is translated into systems and operations), its execution (how the systems and operations are run in production) or its evaluation (how the outputs of the step are assessed to certify the adequacy of the design and its implementation). Finally, as described in the QAF, quality is multi-dimensional.

With consultation with the Policy Committee, the mandate was set to review the quality assurance practices employed during production to assure the release of accurate data. In terms of the aspects described above, the review was to focus on the accuracy in the execution of the steps of the programs. Neither the designs of the programs, nor other dimensions of quality, such as timeliness, were questioned. Furthermore, it was also suggested to focus on the latter steps of the programs where the data are ultimately analyzed and checked prior to data release, as well as the data release step itself. In response to this, the Steering Committee extended the review to the Dissemination and Communications program, thus adding a tenth program.

3. Implementation of the Review

With this mandate in hand, the Steering Committee set out to recruit the employees and develop the material needed to carry out the review.

Ten managers, primarily at the Assistant Director level, were recruited from across the Agency to form the review teams. One team was formed for each of the nine statistical programs listed above, with a separate team formed to review the data release activities of the Dissemination and Communications program (DC). With the exception of the latter team, each team consisted of three managers, so that each manager was involved in the review of three different programs. This interlocking assignment of reviewers to review teams was designed specifically to expose each reviewer to several programs and thereby achieve a more uniform approach. For the same reason, a member of the Steering Committee was assigned to be a member of each review team. For the DC

review, the team consisted of an Assistant Director and all of the members of the Steering Committee.

A lead reviewer (not one of the Steering Committee members) was assigned for each program and was generally responsible for organizing the review, collating all of the materials and preparing a report on the program. In order to avoid any conflict of interest, care was taken to ensure that the lead reviewer did not work in the program being reviewed. In some cases one of the other team members who were familiar with a specific program was assigned to be a member of that review team, in order to provide insight into the program.

The Steering Committee developed a semi-structured questionnaire to be used in the main interviews with the statistical programs. The questionnaire was not designed to be self-completed by programs, but rather was intended to provide the review teams and the programs with suggested lines of questioning, while not being so structured as to confine the review teams to asking pre-defined questions. The questionnaire first collected information on the various steps in the production and the process flows. For each step, the questionnaire then asked about the various kinds of checks in place, indicators of quality at each step, and the various risk factors for each step. The third section of the questionnaire concentrated on the data certification (for example what data confrontation is done or what internal checks are done) to ascertain the accuracy of statistical outputs. The fourth part of the questionnaire asked about the release process and what checks were in place to ensure that there were no errors introduced at this step. The fifth part asked the program about its experiences (if any) where incorrect information had been released in the past, what the reasons were, and what had been done since then to prevent errors in future. The sixth part of the questionnaire asked about various factors that might cause checks to fail more frequently. Finally, the questionnaire asked about how changes to the program, whether internal or imposed from outside, are managed. The questionnaire was sent to the programs in advance of the first meeting to familiarize them with the content of the review.

The actual review started with an introductory meeting with the Quality Secretariat of the review, the lead reviewer, and the director of the program. The objectives of this meeting were to introduce the review process to the program director and introduce the program to the lead reviewer.

One of the challenges described earlier was to identify basic steps that are present in every statistical program (this excludes the Dissemination and Communications program) to which the managers can speak. Based on the introductory meetings with the program managers, the Steering Committee developed a list of seven standard steps which were relevant to the review and common to all programs. These were:

i. Preparation for certification
ii. Data collection (including creating a frame and selecting a sample, if any)
iii. Editing/transformation of data
iv. Imputation and estimation
v. Certification
vi. Release
vii. Post-release

While the details of these steps varied across statistical programs, they were common to all nine statistical programs and this framework proved to be a very useful way of organizing the remainder of the review.

Following the introductory meeting, the lead reviewer organized a series of meetings between the full review team and the program to cover the remainder of the topics in the questionnaire. The program was typically represented by the Director, Assistant Director and/or production manager. It was often difficult to find appropriate meeting times around the busy production schedules of the monthly and quarterly programs. Consequently, the interviews took about four weeks longer to complete than originally planned.

The Steering Committee developed a standard format for the reports on each program. Following a standard introduction describing the background to the review, each report was to cover the basic steps of the program by briefly describing the step, enumerating the various quality assurance checks and identifying the main factors, if any, that could put the program at risk of making errors. The report was to conclude with a summary, describe other relevant considerations that were not in-scope for the review and finish with Appendices. As a guideline, the Steering Committee suggested to the review teams that each report should be approximately 15 pages, excluding any Appendices.

Once each team had drafted its report, the reports were reviewed with the management of the programs. The purpose of this was not to ensure that the program management agreed with all the findings and recommendations, but that the review teams had reflected accurately what it had been told about the program. Following this process the reports were finalized and passed to the Steering Committee. The Steering Committee reviewed the reports and in some cases asked that further follow-ups or clarification be made.

The final step of the review teams' involvement was an all-day debriefing session, where each team presented a summary of its findings and answered questions from the Steering Committee and the other reviewers. As well, the review process was discussed and suggestions were solicited for improvements to future reviews of this type.

With the individual program reports and the discussions at the all-day debriefing session, the Steering Committee prepared a summary report. The report consolidated the most common issues into 29 recommendations. The summary report was presented to the Chief Statistician and its recommendations were discussed at a meeting of the Policy Committee in March 2007. The Policy Committee prepared a formal response to each recommendation. Finally, the summary report and the disposition of recommendations were translated and released in Statistics Canada's Daily on June 4th.

4. Key findings

The review was commissioned by Statistics Canada's Policy Committee in order to assess the soundness of quality assurance processes for nine programs and to propose improvements where needed. The summary report released on June 4th (Statistics Canada, 2007) describes the principal themes that recur frequently throughout these programs, as well as providing guidance for future reviews of this type. The review produced 29 recommendations. The main ones were highlighted in the report's executive summary and are provided in this section.

The review felt that program managers are well aware of the risks factors for their programs and are managing them well, within the circumstances and resources available to them. Hence, they are also well-positioned to identify the improvements that are needed to reduce the risk of errors. Based on the results of this review, programs should be asked to develop proposals for such improvements.

While the risk of errors is well-managed across all programs reviewed, some would benefit more than others from additional investment to further lower the risk of errors. The Consumer Price Index, Balance of Payments and International Trade programs should receive a higher priority for investment to further reduce the risk of errors. It is noted that as writing the report, funds had already been allocated to strengthen the CPI.

Human resources concerns dominate all other risk factors. Investments should first be targeted at addressing the human resources issues affecting the mission-critical programs, to ensure sufficient, well-trained staff to support them. In particular, the existence of a research and analysis capacity separate from the production operation is a key factor in assuring quality. All mission critical programs should have a strong and explicit research and analysis capacity, independent from the production operation, whose role is to challenge the data and to conduct research into the particular subject matter of the program.

There are numerous "best practices" in all programs that can usefully be shared within the Agency. Many of these best practices are small things, but taken together they add up to an effective quality assurance program. These best practices should be the basis of the training program on quality assurance that is currently being developed.

Proposals to further increase the timeliness of the programs should be regarded with extreme caution, especially if there is a possibility that the time would be taken from the certification or data release steps.

In terms of quality management processes, the review of quality assurance practices has been widely regarded by participants as very useful. Statistics Canada should establish an ongoing program of Quality Assurance Reviews, tied to the Integrated Program Reporting cycle, and based on teams of Assistant Directors reporting to a Steering Committee of Directors General. Furthermore, Statistics Canada should consider developing a formal Quality Incident Response Plan containing standard procedures for dealing with data quality incidents, whether before or after data release. Training would be a part of this initiative.

5. Evaluation of the review process

This section presents an evaluation of the review process itself and makes recommendations for future reviews of this nature. This evaluation is based on discussions with the review teams during the all-day debriefing, a short questionnaire completed by the reviewers, and the impressions of the Steering Committee.

It was important to have clear objectives and a clear mandate to carry out so many reviews in such a short time. The framework used to define the quality review in terms of a set of key parameters (outputs vs. process, steps of a process, aspects of a step and dimensions of quality) was very useful in deciding where best to focus the attention of the review to effectively address the events that brought the Policy Committee to launch it. The framework can also serve to highlight the strengths and limitations of the review.

Having a Steering Committee member on each of the review teams, while time-consuming, was beneficial in three ways. First, it helped to standardize the process, since Steering Committee members could give guidance to the review teams. Second, it gave the review teams increased legitimacy with the programs by having someone at a senior level present for the interviews. Third, it gave direct exposure and feedback on the programs to the Steering Committee members. This is an important aspect of the approach that should be retained.

The ten review team members consisted primarily of Assistant Directors. Given the short amount of time available to organize the review, most of the reviewers were identified and "volunteered" by members of the Steering Committee. Despite this, all of the reviewers were enthusiastic and quickly dedicated themselves to the project. They clearly recognized the importance of the project to the organization and were interested in participating, even though they had to fit this work into their already busy schedules. They expected that this would take a considerable amount of time and they were correct: by the end of the process they estimated that they had each spent eight full working days on the project for which they were the lead reviewer. When one adds the time spent in meetings and reviewing reports for the other two reviews in which they participated, preparation for and participation in the one-day debriefing session, and subsequent editing of the reports, it is not unreasonable to assume that the average commitment approached the equivalent of three weeks full-time.

Despite some initial uncertainty, the review process – essentially peer review teams conducting a structured exploration of a program – was a valuable learning experience. The reviewers had the breadth of experience and depth of knowledge to effectively engage program managers. The fact that the review teams were not immersed in the programs in question meant they had to ask questions as "outsiders". The process was iterative and structured, but with enough flexibility to make it meaningful across a diverse set of programs.

Having a fairly large number of reviewers, each working on separate teams, created an atmosphere of learning together and making adjustments "as we go" that offset most of the drawbacks of conducting the reviews simultaneously. It alleviated the pressure on any one individual reviewer and provided a sense of not being alone on this important project.

The fact that each reviewer saw three programs and that *all* reviewers periodically sat around the same table was tremendously useful in dealing with difficult issues and potential unevenness in evaluation standards. It also meant that reviewers were learning about best practices and risks in other programs. At the same time, by assigning a "lead reviewer" for each program, it was clear who was responsible for preparing the final report on each program.

Finally, all reviewers were very happy to have participated in this project (see Table 1). All of the reviewers found the project interesting and a good learning experience. Despite the extra workload, almost all were interested in repeating the experience and nine out of ten highly would recommend such a project to colleagues. The reviewers also felt that this type of evaluation should be applied to other programs.

Table 1: Responses of the ten reviewers to an evaluation questionnaire

Aspect of project	Not at all	A little	A lot	Totally
Project was interesting	0	0	5	5
Project was good learning experience	0	0	4	6
Have a better knowledge of risks that can affect quality	0	1	6	3
Glad to have participated	0	0	4	6
Interested again in a year or two	0	2	6	2
Would recommend a similar project to colleagues	0	1	7	2
Should apply similar evaluation to other programs	0	1	3	6

Based on these results, the review team highly recommended this type of project as an excellent learning experience for the Assistant Director level. Statistics Canada could build on this exercise by recognizing the intense interest of the Assistant Director level in comparing QA practices across programs, with the intention of importing/exporting them. Although this was an *ad hoc* exercise, it proved beyond a doubt that these managers recognize the benefits of this type of peer review process and will willingly engage in it.

Overall, the programs were very cooperative and took the review seriously. When asked whether the "reviewed program was cooperative", all ten reviewers reported either "a lot" or "totally." As an added indication of such cooperation, at the introductory meeting with the Director of the

program, the discussion often started getting into the actual review and had to be cut off until the full review team was available.

The amount of preparation by the programs did vary. Some programs came very well prepared, with well documented processes, while others were more reactive. Having the introductory meeting with the programs was very useful in equalizing this by letting programs know what was expected of them. For future reviews of this type, it would be useful to spend more time in defining what is expected from programs in terms of documentation. Examples from the individual program reports can serve as a useful guide in this respect.

The initial meeting was also very valuable in collecting enough information about the processes to categorize their activities into the seven standard steps listed in Section 4. This served as an excellent way of standardizing the remaining steps in the review (e.g., interviews, report writing) and could probably be applied to future reviews without change.

One of the limitations of the exercise was that the review teams did not have time to speak to the various infrastructure programs areas involved in supporting these statistical programs. The conclusions of the reports are based primarily on information and documentation obtained during a series of interviews with the management of the statistical programs. The one exception is the support provided by the Dissemination and Communications program. Their role was important enough that it was felt vital to interview them. In future, if time permits, the interviews for a specific program should extend to the infrastructure programs, such as Methodology, Systems Development, Business Register and Collection.

The review was launched in September 2006, the training of reviewers took place in October and the review was completed in February 2007, two months later than initially planned. In hindsight, conducting so many reviews with staff dedicated to do a thorough job in such a short timeframe (four months) was overly optimistic. First, it took more interview meetings than expected. On average, each review took 3.2 meetings (not counting the initial meeting) and the total average time per review was 6.1 hours, rather than one or two meetings of two hours as planned. The second reason was that the programs under review are monthly or quarterly production operations and therefore had certain periods when they were not available to meet with the review teams due to production deadlines. Seven of the ten reviewers reported difficulties in scheduling meetings around the programs' production schedules. The interviews also ran into the pre-Christmas holiday period and in a few cases into January. The majority of reviewers mentioned that the holiday break hindered the momentum of the review and a few felt that it had a serious negative effect on the schedule. Overall, half of the reviewers felt that the project deadlines were at least somewhat unreasonable.

As well as the interviews, the process of finalization of the reports also took somewhat longer than planned. There was no explicit step built into the schedule for the Steering Committee members to review and provide comments on the reports, as there should have been. When the reports were reviewed by the Steering Committee members, they realized that some of the reports needed to be reworked. In some cases too much focus was put on one aspect of the program, and in other cases the authors had not provided the necessary arguments to support their conclusions and recommendations. In future reviews, approximately three weeks should be built into the schedule

for such a review. Some good examples of reports were identified and could serve as examples for future reviews.

Table 2 shows a breakdown of the average number of hours spent by each reviewer on their program. The time estimates are approximate based on self-reporting by the reviewers, so they should probably be interpreted only in relative terms. Somewhat surprisingly, additional research on the program took almost twice as long as the interviews themselves; the lead reviewers also spent an average of almost 11 hours researching and consulting other materials to complete their reports. The most relevant documents were those provided by the program, the Quadrennial or Biennial Program Reports, the Integrated Metadata Base record and official releases by the program. They also had to contact the programs on occasion to clarify points discussed in the interviews. As expected, drafting of the report was the most time-consuming activity.

Table 2: Time taken for selected major activities (hours per lead reviewer)

Activity	Hours
Interviews with programs (excluding introductory meeting)	6.1
Conducting additional research on the program	10.9
Writing first draft of report	21.8
Vetting and finalization of report	4.6
Review and commenting on other reports	6.9

Overall, it took four months to assess the quality of 10 major programs, from the time the reviewers and the statistical programs had the instructions and the questionnaire to the completion of the individual and summary reports. With the knowledge and experience gained, it is fair to expect that such a review on as many programs could be completed a few weeks sooner.

The Steering Committee developed a number of documents for the review process. The purpose of these was to standardize the review process as much as possible across the nine statistical programs plus the DC program, while maintaining sufficient flexibility for the review teams to adapt their review to the specifics of the programs.

The main materials were an initial briefing to the review teams, the questionnaire, the report template, and an agenda and instructions for the one-day debriefing session. The briefing seemed to work well, as nine out of ten of the reviewers found the mandate clear. Our impression is that the one-day debriefing session also went well and was appreciated as a good way to "wrap-up" the reviewers' involvement.
The questionnaire was designed as a semi-structured document that was designed to suggest lines of questioning or prompts during the interviews. It was generally successful, although it became clear during the interviews that it contained much more detail than could be covered in a few meetings. The reviewers thus had to do considerable additional research to complete their reports. Nevertheless, all of the reviewers reported that the materials provided were "a lot" or "totally" useful, and nine of them said the same about the materials being sufficient.

The report template was designed to standardize the format of the reports. It provided a general outline, but did not specify particular charts or tables that should be included. While several of the

reports did follow the format, others did not. The template came somewhat late in the process, after some of the reviewers may have already started to formulate the outline of their reports. In other cases (e.g., DC) the content simply did not lend itself to the format had been developed. A number of the reviewers came up with excellent ways of summarizing information in the form of graphics and summary tables. Future reviews should examine these reports and try to develop a more detailed template that would improve the standardization of the reports.

6. Conclusion

The review of quality assurance practices in nine of Statistics Canada's key indicator programs, as well as its Dissemination and Communications program, was developed and carried out very quickly and very successfully thanks to the dedication of program reviewers and the strong cooperation of program managers.

Although the review was just completed prior to drafting this paper, it has already led to some concrete actions and the allocation of additional resources to address the more urgent issues that it raised. The review's focus on a particular aspect of a statistical program, that being on the accuracy in the execution of the program, rather than a broader or more extensive assessment was another key factor in its success. In that sense, the review's focus has proven to be a useful complement to the more extensive assessment of the statistical program outputs that is carried every four years.

In the coming months, Statistics Canada will be building on this positive experience and pursuing its research on developing a formal review process that would be applied to other programs and focusing on other aspects of statistical programs and quality. The objective would be to develop a process that is adaptable to different programs and different needs of the organization. Such a process could take on the form of a tool box of evaluation strategies and associated forms used at Statistics Canada or other statistical organizations.

References

Brackstone, G., (1999). "Managing Data Quality in a Statistical Agency", Survey Methodology, Statistics Canada, 139-149.

Bushery, J. M. McGovern P. D., Marck, P. S., Hogan, H. R. (2006). "Quality Audits at the U.S. Census Bureau", Proceedings of the European Conference on Quality in Survey Statistics (Q2006).

Eiderbrandt-Nilsson, G., (2004). "Quality Audits at Statistics Sweden 2002-2003", Proceedings of the European Conference on Quality and Methodology in Official Statistics (Q2004).

Julien, C. and Born, A., (2006). "Quality Management Assessment at Statistics Canada", Proceedings of the European Conference on Quality in Survey Statistics (Q2006).

Laiho, J. and Nimmergut, A., (2004). "Using Self-Assessments for Data Quality Management – DESAP Experiences", Proceedings of the European Conference on Quality and Methodology in Official Statistics (Q2004).

Statistics Canada (2002), *Quality Assurance Framework*, No. 12-586-XIE, http://www.statcan.ca/english/freepub/12-586-XIE/12-586-XIE2002001.pdf

Statistics Canada (2007), The Daily, Monday, June 4, 2007. "Review of quality assurance practices", http://www.statcan.ca/Daily/English/070604/b070604a.htm

Examen de la qualité des indicateurs clés à Statistique Canada

Claude Julien, Don Royce

Résumé

En septembre 2006, Statistique Canada a créé un comité afin d'effectuer un examen des pratiques en matière d'assurance de la qualité servant à produire neuf indicateurs clés (par exemple, l'emploi, les prix à la consommation et le produit intérieur brut). L'examen a porté sur une foule d'enquêtes auprès des entreprises, d'enquêtes économiques et d'enquêtes auprès des ménages tirées de données sur les répondants, de données administratives ou d'autres produits d'enquête (c.-à-d. des statistiques dérivées) et s'est terminé en mars 2007. Le présent document décrit le processus d'examen et en résume les principales constatations. Il fait également ressortir les avantages de l'examen pour les programmes statistiques, les personnes qui y ont pris part et le Cadre d'assurance de la qualité de Statistique Canada.

Mots clés : examen de la qualité, assurance de la qualité

1. Introduction

Statistique Canada a de longs antécédents de solides pratiques de gestion de la qualité, notamment plusieurs politiques, normes et lignes directrices qui définissent comment l'organisme communique avec les utilisateurs, planifie ses ressources, conçoit et effectue ses enquêtes, diffuse ses données et évalue ses activités. En préparation à une vérification sur la qualité réalisée par le Bureau du vérificateur général (BVG) en 1997, Statistique Canada a réuni ses pratiques de gestion de la qualité en un Cadre d'assurance de la qualité (CAQ) (Statistique Canada, 2002, et Brackstone, 1999) qui définit la qualité des données selon six dimensions (pertinence, exactitude, actualité, cohérence, intelligibilité et accessibilité) et décrit comment chacune d'elles est gérée.

Le cadre a servi à effectuer une autoévaluation de quatre programmes en vue de la vérification de la qualité. Le BVG a été très satisfait des autoévaluations et a recommandé qu'elles soient appliquées à plus grande échelle. Statistique Canada a depuis lors mis en place un mécanisme officiel de production de rapports de programmes intégrés (RPI) selon lequel chaque programme rend compte de la qualité, des coûts, de la gestion des ressources humaines et des orientations stratégiques. Un examen approfondi de chaque programme est effectué tous les quatre ans (rapport quadriennal) et un examen de suivi plus court est réalisé entre deux examens approfondis (rapport biennal). L'examen prend surtout la forme d'une autoévaluation mais un sondage sur la satisfaction des utilisateurs ou un autre genre de consultation semblable est également nécessaire.

Les RPI s'attardent à l'évaluation des produits d'un programme statistique. Au cours de la dernière année, Statistique Canada a effectué des recherches afin d'ajouter un processus d'examen à son système de production de rapports qui porterait sur les processus des programmes plutôt que sur leurs produits. Des stratégies d'examen dans d'autres organisations, comme les vérifications de la qualité de Statistics Sweden (Eiderbrandt-Nilsson, 2004), les vérifications du U.S. Census Bureau

(Bushery, McGovern, Marck et Hogan, 2006) et l'élaboration d'un programme d'autoévaluation (DESAP, Laiho et Nimmergut, 2004) ont été envisagées. Julien et Born (2006) ont présenté cette recherche lors du « European Conference on Quality is Survey Statistics - Q2006 ».

Cette recherche a été brusquement accélérée, en septembre 2006, lorsqu'en réaction à un petit nombre d'erreurs malencontreuses dans les données publiées par Statistique Canada au cours des 18 mois précédents, l'organisme a décidé d'entreprendre un examen des pratiques d'assurance de la qualité de ses programmes statistiques. La deuxième partie du présent document porte sur certains des défis à relever dans la conception de l'examen et les principales décisions qui ont dû être prises. La troisième partie décrit comment l'examen a été instauré. La quatrième partie en présente les principales constatations. La cinquième partie indique les leçons tirées du processus d'examen. La sixième partie fait une conclusion en présentant les recherches futures à faire sur l'élaboration d'un processus d'examen continu qui pourrait être adapté à divers programmes et besoins de l'organisation.

2. Conception de l'examen

Le Comité des politiques de Statistique Canada, composé du statisticien en chef et des statisticiens en chef adjoints, a déterminé neuf programmes à examiner. Il s'agit de six programmes mensuels : l'Enquête sur la population active (EPA), l'Enquête mensuelle sur le commerce de détail (EMCD), l'Enquête mensuelle sur les industries manufacturières (EMIM), l'Indice des prix à la consommation (IPC), le commerce international (CI) et le Produit intérieur brut par industrie (PIB), et de trois programmes trimestriels : les comptes des revenus et dépenses (CRD), la productivité du travail (PT) et la Balance des paiements (BDP). Ces programmes ont été choisis non pas parce qu'ils semblaient être sujets à faire plus d'erreurs, mais parce qu'ils représentent les programmes infra-annuels les plus « essentiels au mandat » de Statistique Canada puisqu'ils publient des données courantes sur la conjoncture économique. La qualité de ces « indicateurs clés » est donc de toute première importance pour une vaste gamme d'utilisateurs.

Pour superviser l'examen, on a créé un comité directeur composé de quatre directeurs généraux, soit de la Direction de la méthodologie, de la Direction de la statistique économique, de la Direction du système de comptabilité nationale et de la Direction des enquêtes auprès des ménages et sur le travail. Le comité directeur était présidé par le directeur général de la méthodologie qui faisait appel au Secrétariat de la qualité, également situé dans la Direction de la méthodologie, pour obtenir de l'aide pendant l'élaboration et la coordination de l'examen.

Sous la direction du Comité des politiques, le comité directeur a établi ses objectifs et une méthode pour examiner ces programmes dans un délai très serré. Selon les directives, les deux objectifs précis de l'examen étaient les suivants :

i. repérer les points faibles manifestes où des actions sont nécessaires ainsi que les facteurs sous-jacents;
ii. repérer les « pratiques exemplaires » qui devraient être adoptées par d'autres programmes.

Cette démarche équilibrée, c'est-à-dire l'examen des aspects positifs et négatifs, a été un principe fondamental de l'examen. L'identification des facteurs sous-jacents a été capitale pour permettre à la haute direction d'affecter les ressources de manière optimale afin de réagir aux problèmes de qualité soulevés pendant l'examen.

L'examen devait être terminé à la fin de l'année civile, c'est-à-dire dans un délai de quatre mois. Étant donné ce délai très court, on a vite décidé d'effectuer les examens de façon simultanée. Cette décision a entraîné deux défis considérables dans le domaine de la conception : élaborer un seul processus d'examen qui pourrait être appliqué rapidement et de manière efficace à chaque programme et définir clairement la portée de l'examen, c'est-à-dire les aspects du programme et les dimensions de la qualité à examiner.

Le premier défi lié à la conception a découlé du grand nombre de caractéristiques des programmes statistiques. L'EPA, l'EMIM et l'EMCD sont des enquêtes « traditionnelles » comprenant une sélection d'unités d'échantillonnage et la collecte de données auprès des répondants; or, la première est une enquête auprès des ménages et les deux autres sont des enquêtes auprès des entreprises. L'IPC est une enquête moins traditionnelle que les trois premières puisque l'information est recueillie par le personnel sur le terrain et d'autres sources. Le CI est une enquête fondée uniquement sur un volume extrêmement élevé de données administratives. Le PIB, les CRD et la PT sont des programmes de statistiques dérivées. Ils n'effectuent pas d'enquêtes par eux-mêmes et recueillent des données statistiques d'autres programmes qu'ils transforment et analysent afin de produire un ensemble de statistiques cohérent et complet. Enfin, la BDP intègre des caractéristiques de tous ces types de programmes. En raison de cette variété, il a été difficile d'élaborer un processus commun afin d'examiner chaque programme.

La portée d'un examen de la qualité est pratiquement illimitée. Étant donné le court délai de l'examen, il était essentiel d'exposer tous les aspects des programmes et de la qualité qui pourraient être examinés. Premièrement, il était évident, en raison du mandat initial donné par le Comité des politiques, que l'examen devait porter sur les processus des programmes plutôt que sur leurs produits. Tel que susmentionné, les produits des programmes sont déjà évalués par les rapports de programmes intégrés. Deuxièmement, le processus d'un programme statistique se compose d'une série d'étapes (p. ex., établir des objectifs, définir une population, créer une base de sondage et sélectionner un échantillon). Troisièmement, on peut examiner diverses dimensions d'une étape particulière : sa conception (comment l'étape est exposée sur papier), sa mise en œuvre (comment la conception se traduit en systèmes et opérations), son exécution (comment les systèmes et opérations sont exécutés dans la production) ou son évaluation (comment les produits d'une étape sont évalués afin d'attester de l'à-propos de la conception et de sa mise en œuvre). Enfin, comme le décrit le CAQ, la qualité est pluridimensionnelle.

En consultation avec le Comité des politiques, le mandat a été établi afin d'examiner les pratiques d'assurance de la qualité employées pendant la production pour garantir la diffusion de données exactes. Pour ce qui est des aspects décrits plus haut, l'examen devait porter sur l'exactitude de l'exécution des étapes des programmes. Ni la conception des programmes ni d'autres dimensions de la qualité, comme l'actualité, n'a été remise en question. En outre, on a proposé de s'attarder aux dernières étapes des programmes où les données sont analysées en bout de ligne et vérifiées avant la diffusion des données, ainsi que l'étape de la diffusion des données. En réaction à cette

proposition, le comité directeur a élargi l'examen au programme de la diffusion et des communications, ajoutant ainsi un dixième programme.

3. Mise en œuvre de l'examen

Connaissant son mandat, le comité directeur s'est proposé de recruter des employés et d'élaborer le matériel nécessaire pour effectuer l'examen.

Dix gestionnaires, surtout au niveau de directeur adjoint, ont été recrutés dans l'ensemble de l'organisme pour former les équipes d'examen. Une équipe a été créée pour chacun des neuf programmes énumérés ci-dessus, et une autre équipe pour examiner les activités de publication des données du programme de la diffusion et des communications (DC). À l'exception de cette dernière équipe, chaque équipe était composée de trois gestionnaires de sorte que chacun a participé à l'examen de trois programmes. Les examinateurs ont été répartis entre les équipes afin que chacun puisse prendre connaissance de plusieurs programmes et adopter ainsi une démarche plus uniforme. Pour la même raison, un membre du comité directeur a été affecté à chaque équipe d'examen. Pour l'examen du DC, l'équipe était composée d'un directeur adjoint et de tous les membres du comité directeur.

Un examinateur en chef (qui n'était pas membre du comité directeur) était affecté à chaque programme et était généralement chargé d'organiser l'examen, de recueillir tous les documents et de rédiger un rapport sur le programme. Pour éviter les conflits d'intérêts, nous avons fait en sorte que l'examinateur en chef ne travaille pas au programme faisant l'objet de l'examen. Dans certains cas, l'un des autres membres de l'équipe qui connaissait bien le programme a été affecté à cette équipe afin qu'il puisse mieux expliquer le programme.

Le comité directeur a élaboré un questionnaire semi-structuré devant servir aux interviews principales avec les programmes statistiques. Il ne s'agissait pas que chacun des programmes remplisse lui-même le questionnaire, mais plutôt que les équipes d'examen et les programmes disposent d'un canevas de questions qui n'étaient pas trop structurées pour éviter que les équipes soient contraintes par des questions définies d'avance. Dans la première partie du questionnaire, on a recueilli des renseignements sur les diverses étapes de la production et du processus. Pour chaque étape, on a ensuite posé des questions sur les divers types de contrôles en place, les indicateurs de qualité et les divers facteurs de risque de chaque étape. La troisième partie du questionnaire portait essentiellement sur l'attestation des données (par exemple, les données sont-elles mises en question et quels sont les contrôles internes) pour valider l'exactitude des données. La quatrième partie portait sur le processus de diffusion et sur les contrôles en place pour faire en sorte qu'il n'y ait pas d'erreurs à cette étape. La cinquième partie portait sur les expériences (le cas échéant) où un programme avait diffusé des données inexactes par le passé, quelles en étaient les raisons et ce qui avait été fait depuis pour prévenir que de telles erreurs ne se répètent. La sixième partie portait sur les divers facteurs pouvant entraîner l'échec plus fréquent des contrôles. Enfin, on a demandé aux programmes comment les changements apportés, soit à l'interne, soit imposés de l'extérieur, étaient gérés. Le questionnaire a été envoyé aux programmes avant la première réunion pour qu'ils puissent prendre connaissance de la teneur de l'examen.

L'examen en soi a débuté par une réunion préliminaire avec le Secrétariat à la qualité, l'examinateur en chef et le directeur du programme. Les objectifs de cette réunion étaient de présenter le processus d'examen au directeur du programme et de présenter le programme à l'examinateur en chef.

L'un des défis décrits plus haut consistait à déterminer les étapes de base communes à chaque programme statistique (à l'exclusion du programme de la diffusion et des communications) à propos desquelles les gestionnaires peuvent parler. En se fondant sur les résultats des premières réunions avec les gestionnaires de programme, le comité directeur a dressé une liste de sept étapes standard qui étaient pertinentes pour l'examen et communes à tous les programmes. Les voici :

i. Préparation à l'attestation
ii. Collecte des données (y compris la création d'une base de sondage et la sélection d'un échantillon, le cas échéant)
iii. Vérification/transformation des données
iv. Imputation et estimation
v. Attestation
vi. Diffusion
vii. Postdiffusion

Ces étapes n'étaient pas nécessairement identiques pour tous les programmes, mais ces activités étaient communes aux neuf programmes et ce cadre nous a paru très utile pour organiser le reste de l'examen.

Après la réunion préliminaire, l'examinateur en chef a organisé une série de réunions entre toute l'équipe d'examen et le programme pour examiner le reste des points du questionnaire. Le programme était généralement représenté par le directeur, le directeur adjoint et/ou le gestionnaire de la production. Il était souvent difficile de trouver de bons moments pour se rencontrer étant donné les calendriers de production chargés des programmes mensuels et trimestriels. Par conséquent, les interviews ont nécessité quatre semaines de plus que ce qu'on avait prévu au départ.

Le comité directeur a élaboré un format standard pour les rapports de chaque programme. Après la présentation standard décrivant le contexte de l'examen, chaque rapport devait porter sur les étapes fondamentales du programme en décrivant brièvement l'étape, en énumérant les diverses vérifications d'assurance de la qualité et en indiquant les principaux facteurs, le cas échéant, qui pourraient faire en sorte que le programme risque de commettre des erreurs. Le rapport devait se terminer par un résumé, en décrivant d'autres questions pertinentes qui ne relevaient pas de l'examen et finir par des annexes. À titre indicatif, le comité directeur a dit aux équipes d'examen que chaque rapport devait avoir environ 15 pages, à l'exclusion des annexes.

Une fois que chaque équipe en eut rédigé une version préliminaire, les rapports ont été passés en revue avec la direction des programmes. Il ne s'agissait pas de s'assurer que la direction du programme acceptait toutes les constatations et recommandations, mais plutôt que les équipes d'examen avaient bien compris ce qu'on leur avait dit au sujet du programme. Ensuite, les rapports ont été mis au point et transmis au comité directeur. Le comité directeur les a passés en revue et, dans certains cas, a demandé que d'autres mesures de suivi ou des éclaircissements soient faits.

Enfin, la dernière fonction des équipes d'examen a été leur participation à une séance de compte rendu d'une journée entière, où chaque équipe a présenté un résumé de ses constatations et a répondu aux questions du comité directeur et des autres examinateurs. En outre, on a discuté du processus d'examen et demandé des propositions d'amélioration en vue d'autres examens du genre.

Grâce aux rapports de chaque programme et aux discussions lors de la séance de compte rendu d'une journée, le comité directeur a préparé un rapport sommaire qui regroupait les questions les plus courantes en 29 recommandations. Ce rapport a été présenté au statisticien en chef et les recommandations qu'il comprenait ont fait l'objet de discussions lors d'une réunion du Comité des politiques, en mars 2007. Le Comité des politiques a préparé une réponse officielle à chaque recommandation. Enfin, le rapport sommaire et la disposition des recommandations ont été traduits et diffusés le 4 juin dans le Quotidien de Statistique Canada.

4. Principales constatations

L'examen a été demandé par le Comité des politiques de Statistique Canada en vue d'évaluer la validité des processus d'assurance de la qualité de neuf programmes et de proposer les améliorations nécessaires. Le rapport sommaire diffusé le 4 juin (Statistique Canada, 2007) décrit les thèmes principaux qui reviennent fréquemment dans tous ces programmes et contient des recommandations en vue d'autres examens du genre. L'examen a donné 29 recommandations. Les principales sont soulignées dans le résumé du rapport et apparaissent dans la présente partie.

L'examen a permis de déterminer que les gestionnaires de programme connaissaient bien les facteurs de risque que présentaient leurs programmes et qu'ils les géraient bien, compte tenu des circonstances et des ressources à leur disposition. Ils sont bien placés pour déterminer quelles améliorations sont nécessaires afin de réduire le risque d'erreurs. Étant donné les résultats du présent examen, on devrait demander aux programmes d'élaborer des propositions d'améliorations.

Bien que le risque d'erreurs soit bien géré dans l'ensemble des programmes examinés, certains bénéficieraient plus que d'autres d'un investissement supplémentaire afin d'atténuer ce risque. Il faudrait que les programmes de l'Indice des prix à la consommation, de la Balance des paiements et du Commerce international aient une priorité plus élevée en matière d'investissement afin de réduire davantage le risque d'erreurs. Nous constatons, au moment de la rédaction du présent rapport, que des fonds ont été attribués au renforcement de l'IPC.

Les problèmes de ressources humaines dominent tous les autres facteurs de risque. L'investissement devrait cibler en premier lieu les problèmes de ressources humaines touchant les programmes essentiels au mandat de Statistique Canada afin qu'ils puissent disposer d'un personnel suffisant et bien formé. L'existence d'une capacité de recherche et d'analyse distincte de l'activité de production est un facteur essentiel à l'assurance de la qualité. Tous les programmes essentiels au mandat doivent avoir une capacité de recherche et d'analyse solide et explicite, indépendante de l'activité de production, dont le rôle consiste à mettre en question les données et à faire des recherches dans le domaine de spécialité du programme.

De nombreuses « pratiques exemplaires » dans tous les programmes peuvent être utilement mises en commun. Beaucoup de ces pratiques sont modestes, mais ensemble, elles constituent un bon programme d'assurance de la qualité. Ces pratiques exemplaires devraient être le fondement du programme de formation sur l'assurance de la qualité qui est actuellement en voie d'élaboration.

Les propositions visant à augmenter l'actualité des programmes doivent être examinées avec la plus grande prudence, surtout s'il y a une possibilité que du temps soit soustrait aux étapes de l'attestation ou de la diffusion des données.

Pour ce qui est des processus de gestion de la qualité, les participants ont tous jugé très utile l'examen des pratiques d'assurance de la qualité. Statistique Canada devrait établir un programme permanent d'examen de l'assurance de la qualité lié aux cycles des rapports de programme intégrés, et fondé sur des équipes composées de directeurs adjoints relevant d'un comité directeur composé de directeurs généraux. En outre, Statistique Canada devrait songer à élaborer un plan officiel de réaction aux incidents relatifs à la qualité contenant des procédures standard pour régler les incidents liés à la qualité des données, avant ou après leur diffusion. La formation devrait faire partie de cette initiative.

5. Évaluation du processus d'examen

La présente partie contient une évaluation du processus d'examen et des recommandations en vue des prochains examens de cette nature. L'évaluation est fondée sur les discussions avec les équipes d'examen au cours de la séance de compte rendu d'une journée entière, sur un questionnaire rempli par les examinateurs et sur les impressions du comité directeur.

Il est important d'avoir des objectifs et un mandat clairs pour effectuer un si grand nombre d'examens en un si cours laps de temps. Le cadre de travail employé pour définir l'examen de la qualité à l'aide d'un ensemble de paramètres clés (produits et processus, étapes d'un processus, aspects d'une étape et dimensions de la qualité) a été très utile pour décider où il valait mieux centrer l'examen pour réagir comme il se doit aux événements qui ont porté le Comité des politiques à l'instaurer. Le cadre peut également servir à souligner les points forts et les limites de l'examen.

Il a été avantageux sur trois plans, même si cela a pris beaucoup de temps, de faire participer un membre du comité directeur à chacune des équipes d'examen. En premier lieu, il a été plus facile d'uniformiser le processus, puisque les membres du comité directeur ont pu guider les équipes d'examen. En deuxième lieu, les équipes d'examen ont eu ainsi une meilleure légitimité auprès des programmes puisqu'une personne d'un niveau assez élevé était présente aux interviews. En troisième lieu, les membres du comité directeur ont pu prendre directement connaissance des caractéristiques des programmes et des commentaires sur chacun. Il s'agit d'un aspect important de la démarche à retenir.

Les membres des dix équipes d'examen étaient surtout des directeurs adjoints. Étant donné le peu de temps disponible pour organiser l'examen, la plupart des examinateurs ont été identifiés et déclarés « volontaires » par les membres du comité directeur. Malgré cela, tous les examinateurs ont manifesté de l'enthousiasme et se sont rapidement mis à la tâche. Ils savaient à quel point le

projet était important pour l'organisation et ils voulaient y participer, même s'ils devaient intégrer ce travail dans leur horaire déjà chargé. Ils s'attendaient à ce que ce travail prenne beaucoup de temps et ils n'ont pas eu tort : à la fin du processus, ils ont calculé qu'ils avaient consacré chacun huit jours ouvrables complets au projet dont ils étaient l'examinateur en chef. Si on ajoute le temps consacré aux réunions et à la lecture des rapports pour les deux autres examens auxquels ils ont participé, la préparation et la participation à la séance de compte rendu d'une journée et à la révision ultérieure des rapports, il n'est pas déraisonnable de supposer que chacun a consacré au projet l'équivalent de trois semaines à plein temps en moyenne.

Malgré une certaine incertitude au début, le processus d'examen – essentiellement des équipes d'examen composées de pairs qui dirigent de façon structurée l'exploration d'un programme – a été une formidable expérience d'apprentissage. Les examinateurs avaient toute l'expérience et toute la profondeur de connaissances nécessaires pour réussir à mobiliser les gestionnaires de programme. Comme les membres des équipes d'examen n'étaient pas partie prenante des programmes en question, ils ont dû poser des questions comme s'ils étaient des « profanes ». Le processus a été itératif et structuré, mais il a été assez souple pour pouvoir être utile dans un ensemble diversifié de programmes.

Étant donné le nombre assez imposant d'examinateurs, chacun affecté à une équipe distincte, il s'est établi un climat d'apprentissage collectif et des rajustements ont été faits au fur et à mesure de sorte que la plupart des inconvénients liés à la tenue simultanée des examens ont été annulés. C'est ainsi que la pression a été atténuée sur chacun des examinateurs qui avait ainsi le sentiment de ne pas être seul à travailler à cet important projet.

Comme chacun des examinateurs a pu prendre connaissance de trois programmes et que *tous* les examinateurs se sont regroupés périodiquement autour d'une même table, il a été beaucoup plus facile de régler des problèmes difficiles et d'assurer l'uniformité des normes d'évaluation. De cette façon, les examinateurs ont aussi pu connaître les pratiques exemplaires et les risques dans d'autres programmes. Parallèlement, comme un « examinateur en chef » avait été désigné pour chaque programme, on savait clairement qui était chargé de rédiger le rapport final sur chaque programme.

Enfin, tous les examinateurs ont été très heureux de leur participation à ce projet (voir le tableau 1). Tous ont trouvé que le projet a été intéressant et qu'il a été une bonne expérience d'apprentissage. Malgré la surcharge de travail, tous aimeraient répéter l'expérience et neuf sur dix recommanderaient fortement un tel projet à leurs collègues. Ils ont aussi mentionné que d'autres programmes devraient faire l'objet d'une telle évaluation.

Tableau 1 : Réponses des dix examinateurs à un questionnaire d'évaluation

Aspect du projet	Pas du tout	Un peu	Beaucoup	Totalement
Le projet a été intéressant	0	0	5	5
Le projet a été une bonne expérience d'apprentissage	0	0	4	6
J'ai une meilleure connaissance des risques qui peuvent avoir un effet sur la qualité	0	1	6	3
Je suis heureux (heureuse) d'y avoir participé	0	0	4	6
Je serais de nouveau intéressé(e) dans une année ou deux	0	2	6	2
Je recommanderais un projet semblable aux collègues	0	1	7	2
Une évaluation semblable devrait être appliquée à d'autres programmes	0	1	3	6

Étant donné ces résultats, l'équipe d'examen a fortement recommandé ce type de projet comme excellente expérience d'apprentissage au niveau des directeurs adjoints. Statistique Canada pourrait tirer parti de cette activité en reconnaissant le vif intérêt des titulaires d'un poste de directeur adjoint pour la comparaison des pratiques d'AQ entre les programmes, en vue de les importer ou de les exporter. Ce fut dans le présent cas une activité ponctuelle, mais elle a révélé sans aucun doute que ces gestionnaires reconnaissaient les avantages de ce type d'examen par les pairs et qu'ils sont fortement disposés à y participer.

Dans l'ensemble, les programmes ont très bien collaboré et ont pris l'étude au sérieux. Lorsqu'on leur a demandé si le « programme évalué était coopératif », les dix examinateurs ont répondu « beaucoup » ou « totalement ». Autre indice de la collaboration : lors de la réunion préliminaire avec le directeur du programme, la discussion en venait souvent à porter sur l'examen en soi et devait être interrompue jusqu'à ce que toute l'équipe d'examen soit disponible.

Le degré de préparation de chacun des programmes était différent. Certains sont arrivés bien préparés, avec des processus bien documentés, tandis que d'autres ont été plus passifs. Il a été très utile de tenir une réunion préliminaire avec les programmes à cet égard pour les aviser de ce que

l'on attendait d'eux. Pour d'autres examens de ce genre, il serait utile de consacrer plus de temps à la définition de ce qu'on attend des programmes en matière de documentation. On pourrait s'inspirer à cet égard du rapport de certains programmes.

La réunion initiale a aussi été très utile, car elle a permis de recueillir assez de renseignements au sujet des processus pour classer leurs activités dans les sept étapes standard énumérées à la partie 4. Ce fut une excellente façon d'uniformiser les autres étapes de l'examen (p. ex., les interviews et la rédaction des rapports) et on pourrait probablement s'en inspirer sans modification pour les prochains examens.

Il y a eu toutefois certaines contraintes, notamment le fait que les équipes d'examen n'ont pas eu le temps de parler aux divers secteurs de programmes d'infrastructure qui appuient ces programmes statistiques. Les conclusions des rapports reposent surtout sur les renseignements et les documents obtenus au cours d'une série d'interviews avec la direction des programmes statistiques. La seule exception est le soutien offert par le programme de la diffusion et des communications. Son rôle est assez important pour qu'on ait jugé essentiel de l'interviewer. Dans l'avenir, si le temps le permet, les interviews relatives à un programme donné devraient aussi être faites auprès des autres programmes d'infrastructure, comme la Méthodologie, le Développement des systèmes, le Registre des entreprises et la Collecte.

L'examen a débuté en septembre 2006, la formation des examinateurs a eu lieu en octobre et l'examen s'est terminé en février 2007, deux mois plus tard que prévu. En rétrospective, l'exécution d'un si grand nombre d'examens avec le personnel chargé de faire un travail approfondi en un si court laps de temps (quatre mois) était très optimiste. En premier lieu, il a fallu organiser plus de réunions que ce qui avait été prévu. En moyenne, chaque examen a exigé 3,2 réunions (sans compter la réunion initiale) et la durée moyenne totale de chaque examen a été de 6,1 heures, plutôt qu'une réunion ou deux de deux heures comme cela avait été prévu. La deuxième raison, c'est que les programmes examinés ont un calendrier de production mensuel ou trimestriel et qu'ils n'étaient pas disponibles pour rencontrer les équipes d'examen en raison des délais de production. Sept des dix examinateurs ont signalé des difficultés à organiser des réunions, compte tenu du calendrier de production des programmes. Les interviews ont aussi été faites juste avant la période de congé des Fêtes et, dans quelques cas, en janvier. La plupart des examinateurs ont mentionné que le congé des Fêtes a brisé l'élan de l'examen et quelques-uns ont dit qu'il a eu un effet négatif grave sur le calendrier. Dans l'ensemble, la moitié des examinateurs étaient d'avis que l'échéancier du projet était quelque peu déraisonnable.

Tout comme les interviews, la mise au point des rapports a pris plus de temps que prévu. Il n'y avait pas d'étape explicite prévue au calendrier pour l'examen des rapports et la formulation des commentaires par les membres du comité directeur, ce qui aurait dû être le cas. Après que les membres du comité directeur eurent examiné les rapports, ils ont constaté que certains devaient être remaniés. Dans certains cas, on avait accordé trop d'attention à un seul aspect du programme et, dans d'autres cas, les auteurs n'avaient pas avancé les arguments nécessaires pour étayer leurs conclusions et recommandations. Dans les prochains examens, il faudrait prévoir environ trois semaines au calendrier pour cet examen. De bons exemples de rapports ont été identifiés et pourraient servir d'exemples pour les prochains examens.

Le tableau 2 présente une ventilation du nombre moyen d'heures que chaque examinateur a consacré à son programme. Le temps indiqué est approximatif et il est fondé sur l'autodéclaration des examinateurs, de sorte qu'il faut interpréter les données seulement de façon relative. Fait quelque peu étonnant, les recherches supplémentaires sur le programme ont pris presque deux fois plus de temps que les interviews comme telles; les examinateurs en chef ont aussi consacré en moyenne près de 11 heures à lire et à consulter d'autres documents pour mettre au point leurs rapports. Les documents les plus utiles ont été ceux qui avaient été fournis par le programme, les rapports quadriennal ou biennal de programme, le registre de la base de métadonnées intégrées et les communiqués officiels du programme. Ils ont eu aussi à communiquer avec les programmes à l'occasion pour clarifier des points mentionnés pendant les interviews. Comme on s'y attendait, la rédaction du rapport a été l'activité qui a pris le plus de temps.

Tableau 2 : Temps nécessaire à certaines activités principales (heures par examinateur en chef)

Activité	Heures
Interviews avec les programmes (à l'exclusion de la réunion préliminaire)	6,1
Recherches supplémentaires sur le programme	10,9
Rédaction de la première version du rapport	21,8
Examen approfondi et mise au point du rapport	4,6
Examen et commentaires sur d'autres rapports	6,9

Dans l'ensemble, il a fallu quatre mois pour évaluer la qualité de 10 programmes majeurs, depuis le moment où les examinateurs et les programmes statistiques ont obtenu les directives et le questionnaire jusqu'à l'achèvement des rapports individuels et sommaires. Grâce aux connaissances et à l'expérience acquises, il est juste de s'attendre à ce qu'un tel examen sur un aussi grand nombre de programmes puisse prendre quelques semaines de moins.

Le comité directeur a rédigé un certain nombre de documents pour l'examen. Il s'agissait d'uniformiser le processus d'examen le plus possible pour l'ensemble des neuf programmes, plus le programme du DC, tout en accordant aux équipes d'examen assez de latitude pour qu'elles adaptent leur examen aux particularités des programmes.

Les principaux documents ont été ceux de la première séance d'information à l'intention des équipes d'examen, le questionnaire, le modèle de rapport, ainsi qu'un ordre du jour et des instructions en vue de la séance de compte rendu d'une journée. La séance d'information a semblé bien se dérouler puisque neuf des dix examinateurs ont trouvé le mandat clair. Nous avons aussi l'impression que la séance de compte rendu d'une journée s'est bien déroulée et a été appréciée parce qu'il s'agissait d'un bon moyen de faire la récapitulation de la participation des examinateurs.

Le questionnaire était un document semi-structuré conçu pour proposer un canevas de questions ou de questions incitatives au cours des interviews. Il a été généralement utile, mais on s'est rendu compte pendant les interviews qu'il contenait beaucoup plus de matière que ce qui pouvait être couvert en quelques réunions. Les examinateurs ont donc dû faire beaucoup de recherches supplémentaires pour mettre au point leur rapport. Malgré cela, tous les examinateurs ont signalé

que les documents fournis ont été « beaucoup » ou « totalement » utiles et neuf d'entre eux ont dit que les documents avaient été suffisants.

Le modèle de rapport a été conçu afin d'en uniformiser la présentation. Il s'agit d'un synopsis et on n'avait pas mentionné expressément quels graphiques ou tableaux devaient être inclus. Plusieurs rapports ont été présentés conformément au modèle, d'autres pas. Le modèle a été proposé plutôt tard dans le processus, après que certains examinateurs eurent déjà commencé à formuler le synopsis de leur rapport. Dans d'autres cas (p. ex., DC), le contenu ne se prêtait tout simplement pas à la présentation recommandée. Quelques examinateurs ont proposé d'excellentes façons de résumer l'information sous la forme de graphiques et de tableaux sommaires. Les examens à venir devraient examiner ces rapports et tâcher d'élaborer un modèle plus détaillé qui améliorerait l'uniformité des rapports.

6. Conclusion

L'examen des pratiques d'assurance de la qualité dans neuf programmes d'indicateurs clés de Statistique Canada, ainsi que son programme de la diffusion et des communications, a été élaboré et réalisé très rapidement et avec succès grâce au dévouement des examinateurs des programmes et de la grande collaboration des gestionnaires de programmes.

Même si l'examen n'a été achevé que peu avant la rédaction du présent document, il a déjà entraîné certaines mesures concrètes et l'affectation de ressources supplémentaires en réaction aux questions les plus urgentes qui ont été soulevées. Le fait que l'examen porte sur un aspect particulier d'un programme statistique, soit celui de l'exactitude dans l'exécution du programme, plutôt que sur une évaluation plus large ou plus approfondie est un autre facteur clé de sa réussite. En ce sens, la portée de l'examen s'est avérée être un complément utile de l'évaluation plus étendue des produits des programmes statistiques qui est effectuée tous les quatre ans.

Au cours des mois à venir, Statistique Canada s'inspirera de cette expérience positive et poursuivra ses recherches sur l'élaboration d'un processus d'examen officiel qui serait appliqué à d'autres programmes et porterait sur d'autres aspects des programmes statistiques et de la qualité. L'objectif serait d'élaborer un processus qui puisse être adapté à différents programmes et besoins de l'organisation. Un tel processus pourrait prendre la forme d'une trousse d'outils contenant des stratégies d'évaluation et des formulaires connexes employés à Statistique Canada ou dans d'autres organismes statistiques.

Références

Brackstone, G., (1999). "La gestion de la qualité des données dans un bureau de statistique", Techniques d'enquête, Statistique Canada, 159-171.

Bushery, J. M. McGovern P. D., Marck, P. S., Hogan, H. R. (2006). "Quality Audits at the U.S. Census Bureau", Recueil du European Conference on Quality in Survey Statistics (Q2006).

Eiderbrandt-Nilsson, G., (2004). "Quality Audits at Statistics Sweden 2002-2003", Recueil du European Conference on Quality and Methodology in Official Statistics (Q2004).

Julien, C. and Born, A., (2006). "Quality Management Assessment at Statistics Canada", Recueil du European Conference on Quality in Survey Statistics (Q2006).

Laiho, J. and Nimmergut, A., (2004). "Using Self-Assessments for Data Quality Management – DESAP Experiences", Recueil du European Conference on Quality and Methodology in Official Statistics (Q2004).

Statistique Canada (2002), *Le cadre d'assurance de la qualité de Statistique Canada*, No. 12-586-XIF, http://www.statcan.ca/francais/freepub/12-586-XIF/12-586-XIF2002001.pdf

Statistique Canada (2007), Le Quotidien, Lundi, 4 juin 2007. "Examen des pratiques d'assurance de la qualité", http://www.statcan.ca/Daily/Francais/070604/e070604a.htm

STATISTICS CANADA LIBRARY
BIBLIOTHÈQUE STATISTIQUE CANADA
1010437458