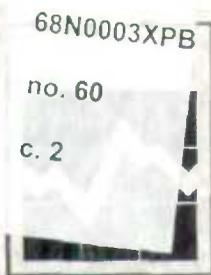


Available also on the StatCan INTRANET site /  
Disponible aussi sur le site INTRANET de StatCan



**Project to Improve Provincial Economic  
Statistics**

**Projet d'amélioration des statistiques  
économiques provinciales**

## **Coherence Analysis of Large Enterprise Data**

## **Analyse de la cohérence des données sur les grandes entreprises**

Technical Series

Série technique

Number 60

Numéro 60



Internet: [www.statcan.ca](http://www.statcan.ca)  
Intranet: <http://pipes>



Statistics  
Canada

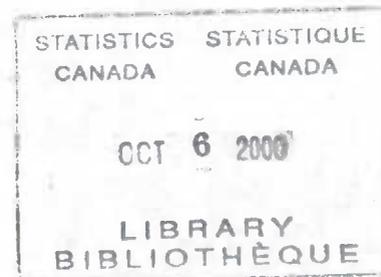
Statistique  
Canada

**Canada**

# Coherence Analysis of Large Enterprise Data

Large enterprises have a significant impact on the Canadian economy and are consequently likely to be included in most of Statistics Canada's (STC) various business surveys. Incomplete and/or inaccurate reporting by these enterprises can substantially impact the quality of the estimates produced by the agency. It is therefore important to ensure that the survey information provided for individual business units within an enterprise is consistent with the consolidated income statement and balance sheet data provided for the enterprise as a whole.

Over the last number of years Statistics Canada has been analyzing the coherence of its business statistics. This analysis has ranged from case studies of large enterprises to analysis of head office and ancillary unit data as well as smaller complex enterprises. This paper presents the reader with the highlights of these analyses, the main impediments to producing coherent data and what measures and processes a statistical agency can adopt to move toward producing more coherent data.



## Note of appreciation

*Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.*

For further information on the materials covered in this paper, please contact Bonnie Bercik (613) 951-6790 or Diane Proulx (613) 951-7192  
Fax: (613) 951-0411

# Coherence Analysis of Large Enterprise Data

This paper was presented at the Second International Conference on Establishment Surveys (ICES - II) in Buffalo, New York, on June 19, 2000.

Monique Gaudreau, Jim Tebrake, Statistics Canada  
Statistics Canada, Main Building, Room 1105, Ottawa, Ontario Canada K1A 0T6  
[gaudmon@statcan.ca](mailto:gaudmon@statcan.ca) [tebrajam@statcan.ca](mailto:tebrajam@statcan.ca)

**Key words:** coherence analysis, incoherence, complex enterprise, Project to Improve Provincial Economic Statistics (PIPES), Unified Enterprise Survey, Key Provider Manager.

## 1. Background

In early 1997, Statistics Canada launched a large and important "Project to Improve Provincial Economic Statistics" (PIPES). In broad terms, the goal of PIPES is to bring about a permanent and very substantial increase in the volume and quality of provincial and territorial statistical information. Improvements will be made under four distinct headings that are sometimes referred to as "the four C's" of data quality: coverage, counts, consistency and coherence.

Coherence is defined as a situation where information on what businesses sold (supply-side data) balances with information on what consumers spent (demand-side data). *In addition, data collected from establishments, and data collected from their parent enterprise, must be internally consistent, comprehensive, and not duplicated.*

To implement PIPES, Statistics Canada is redesigning its entire framework for conducting business surveys. Prior to 1997, Statistics Canada conducted some 200 business surveys each year, using about 700 distinct questionnaires. These surveys were the responsibility of individual program managers, organized in a dozen or so subject matter divisions, each responsible for a specific industrial sector such as agriculture, transportation or manufacturing divisions. For several surveys, the production processes were vertically integrated, that is, conducted completely within those subject matter divisions. The production processes included such steps as generating and mailing out questionnaires, data capture, weighting, aggregating and tabulating the data, analyzing the results and preparing information products for dissemination. Some surveys centralized certain steps, such as data capture or questionnaire mailing, in specialized "service" divisions. This alternative is usually more efficient and also has the advantage of promoting greater consistency across surveys or sectors.

Statistics Canada presently promotes the use of consistent concepts, methods and procedures across its business surveys, through a number of mechanisms. These include standard conceptual frameworks (such as the System of National Accounts, SNA), standard classification systems (such as the North American Industrial Classification System, NAICS), a common business register, common staff pools for methodology, operations and systems development, and corporate policies related to survey-taking procedures. Nevertheless, there still exists some degree of inconsistency and content overlap across Statistics Canada's business surveys. The data being collected from business enterprises also suffer from a degree of incoherence.

Beginning in 1997, many of the 100 annual surveys are being phased into a single master survey program called the Unified Enterprise Survey (UES). The new integrated approach focuses on enterprises: it ensures that financial data received from an enterprise's head office, such as corporate profits, are consistent with production and sales data received from its different establishments. The UES collects more industry and commodity detail at the provincial level and avoids overlap between different survey questionnaires.

Statistics Canada established the Key Provider Manager (KPM) program under its new Project to Improve Provincial Economics Statistics (PIPES). The objectives of this program are to enhance Statistics Canada's working relations with large and complex enterprises; to improve response in terms of timeliness, completeness and accuracy; and to increase the quality and consistency of reported data. On this latter point, Key Provider Managers must ensure that the data collected from the enterprises in their portfolios can be treated as a set of coherent and interrelated data. The ability to analyse the entire range of data for particular business enterprises will lead to greater internal data coherence as well as form the basis on which to address other issues such as alternative imputation methods.

## **2. Why is Enterprise Level Coherence Analysis Important?**

Why should a statistical agency be concerned about producing coherent data for individual enterprises? After all, the main goal is to produce high quality industry level/provincial level estimates to be used by the system of national accounts. Whether or not data for a single enterprise are coherent shouldn't be a major concern. Or should it? In this next section we outline in what cases it is appropriate to undertake coherence analysis and the benefits this analysis yields.

In Canada the economic landscape is dominated by a small number of very large firms. Because business activity tends to be skewed in most industries, certain key enterprises have a large impact on the final industry estimates. In those cases where industries are dominated by a few large firms coherence analysis is necessary. Enterprise level coherence analysis is necessary as it produces two major benefits.

The first benefit from enterprise level coherence analysis is the identification of the most serious of errors - double counting. Consider the following scenario where Enterprise A has been receiving survey questionnaires requesting information for their manufacturing and wholesale activities, has been reporting for the past 5 years, has been consistent from year to year but because they cannot provide the requested breakdown has therefore been reporting the same figures for their entire operations on both the manufacturing and wholesale questionnaires.

Without engaging in some form of coherence analysis there is no assurance that double counting will be identified. The result is that estimates of the wholesale industry will include manufacturing activity and estimates for the manufacturing industry will include wholesale activity. As will be seen later by using an enterprise-centric approach to analysis these types of errors will be caught.

The second benefit of enterprise level coherence analysis is the identification of under-coverage. A common challenge of all business surveys is trying to ensure an up-to-date frame. One way to verify coverage is comparing an enterprise level total to the sum of its parts. If the enterprise total is larger it may indicate under-coverage. A good way to illustrate this point is to draw an analogy to the type of editing that goes on in most surveys. Most surveys have a revenue section and in

most cases they ask for both total revenue (control total) and individual revenue items. If the total and the sum of the individual components are not equal imputation is undertaken. Using coherence analysis, the total revenue represents the enterprise total and the individual parts represent the different establishments in the enterprise. If the enterprise total is not compared to the sum of the parts it is impossible to ensure complete coverage of the enterprise.

One might argue that the two benefits mentioned above tend to offset each other, and therefore enterprise level coherence analysis is not really that important. However, these large enterprises are extremely unique and therefore this canceling-out effect is unlikely.

## 2.1 KPM Enterprises

Large enterprises have a significant impact on the Canadian economy and are consequently likely to be included in Statistics Canada's various surveys. It is therefore important to ensure that consistent data are collected both at the financial unit and production unit levels, and accurately reflected in the data on Canada's economy.

A case study was undertaken involving two large corporations in the Key Provider Manager Program. The analysis consisted of comparing the sum of values reported at the establishment level to the value reported at the enterprise level taking into account internal and external sales and transfers. Of the two enterprises selected one was fully contained in the UES, meaning that all of its establishments were in the UES sample, while the other was partially contained, meaning that some establishments were not in the UES sample. Seven key variables were used in the analysis:

1. goods and services expenditures (before the UES);  
total operating expenses (after the UES);
2. revenue from the sale of goods and services (before the UES);  
total operating revenue (after the UES);
3. gross operating profit;
4. number of employees;
5. wages and salaries;
6. capital expenditures;
7. opening and closing inventories.

Officials in the company fully contained within the UES proceeded to do the analysis themselves and correct any inconsistencies they found. It took 60 hours for the corporation to reconcile the data it had already given to us.

The analysis of the non-UES company's data consisted of comparing the sum of values reported at the establishment level to the values reported at the enterprise level taking into account internal and external sales and transfers. The eight variables listed above were analyzed. Differences ranged from as low as 1% in Total Operating Revenue to as high as 44% in inventory between the information obtained from the enterprise and that obtained from the establishments. Reasons for incoherence included differences in concepts and definitions used by the survey areas (prior to the UES), coverage and reporting errors, processing errors and transfers.

In both cases it was not a question of whether or not the data could be reconciled - they could. The question was how much work it took to reconcile the data and what procedures could be put in place to ensure that incoherence would not happen in the future.

The study also resulted in recommendations for a Coherence Analysis Framework for the KPM Program.<sup>1</sup> The recommendations can be summarized as follows:

1. a systematic approach should be developed to gather and present the information;
2. the value added calculation should be automated;
3. administrative data should be part of coherence analysis;
4. acceptance thresholds by variable should be defined.

## 2.2 Beyond the Case Study

Coherence analysis is a new concept. In many ways most analysts have engaged in some form of coherence analysis in the past. An analyst with a large respondent in one survey will often inquire as to what the activity of that same respondent is in another survey. What has been missing, at least in many cases, is that a 'back of a napkin' approach has often been taken when performing this type of analysis. Rarely was anything formal or concrete ever put into place. This section of the paper details the steps needed to move beyond case studies and establish coherence as an integral part of the annual business survey program. Specifically we have identified three processes needed for coherence analysis. These include:

- Reconciliation
- Data Assembly
- Data Analysis

### 2.2.1 Reconciliation

Coherence Analysis is about bringing data together. While in principle this may appear to be only a technical task it is in fact only a small part of the work. Before the data can be brought together three types of reconciliation must occur.

- **Frame Reconciliation**  
One of the main impediments to coherence analysis is that often the same frame is not used for all surveys. In such situations, a manual exercise of linking the data together must then be undertaken. Often this reconciliation exercise is so time-consuming and cumbersome that it is not done or not done in time to be of use.
- **Conceptual Reconciliation**  
Problems also occur when individual surveys ask their questions differently or attempt to measure things in a slightly different manner. The end result is an apples and oranges comparison and the need to make certain assumptions to accurately compare the data.
- **Processing Reconciliation**  
If successful in linking the frames and reconciling the conceptual differences, the analyst must still contend with the differences in the way the data were processed. Was the edit and imputation system based on donor records or historical records? If historical edits were used, was the same growth rate applied across surveys for the same enterprise? Were the data allocated across provinces and if so, was the same allocator used across all surveys.

The Unified Enterprise Survey (UES) program at Statistics Canada has removed the reconciliation problem associated with coherence analysis. Each survey is drawn from the same frame and a central unit helps ensure that the questions asked on each survey are as conceptually similar as possible. The processing of the data has been centralized ensuring that the same methods are used at the capture, edit, imputation and allocation stages. This more than anything has made it possible to perform coherence analysis.

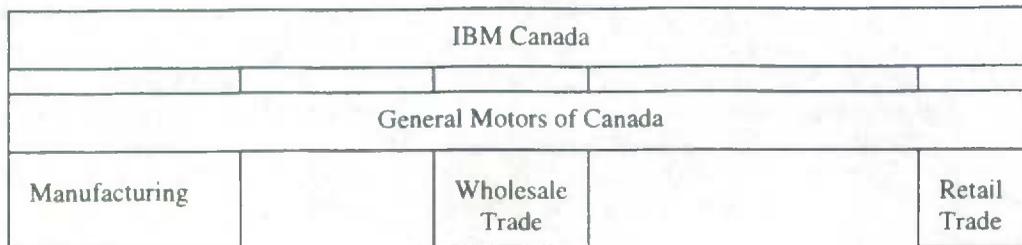
---

<sup>1</sup> See PIPES Technical series Number 31- Coherence Analysis – Case Study from the Key Provider Manager Program

### 2.2.2 Data Assembly

While the changes noted above have contributed substantially to make coherence analysis possible, it is still necessary to assemble the data for the analysis to take place. For the most part data are still assembled by industry. This is a logical choice given that the end goal is to produce reliable industry estimates of value added. As a result, the databases and tools used to extract data are structured by industry. However, to ensure coherence, the data should also be presented from an 'enterprise-centric' viewpoint.

Consider the illustration below. Instead of grouping establishments by industry (e.g. group of wholesale establishments, group of manufacturing establishments) we need to produce databases which group establishments within their overall organization structure. If you will, we need to start analyzing the annual survey of IBM or the annual survey of General Motors. We then need to process these data, as we would process a group of establishments within the same industry. If we do this we will go a long way in producing coherent data.



Currently at Statistics Canada we have many different sources of information for a particular enterprise. As part of the UES project we are assembling an enterprise-centric database which makes this type of analysis possible for the different enterprises. For any particular enterprise all or some of the following information exists:

1. Enterprise Frame (Central Business Register)
2. Enterprise data from enterprise-based survey questionnaires
3. Establishment Frame (Central Business Register)
4. Establishment data from establishment-based survey questionnaires in the UES
5. Legal Financial Data (Corporate Income Tax returns)
6. Wages and Salaries (Payroll Deduction Administrative Data)

While an analyst must fully understand what each of these sources of information is attempting to cover they must be compared against each other in order to paint a picture of the information a particular enterprise is sending to Statistics Canada. The first task of coherence analysis is then to simply assemble the data such that a picture can be drawn for the analyst. Below is an example of such a picture.

#### Enterprise A

Source	Operating Revenue ( '000)	Operating Expenses ( '000)
Enterprise total	475,499	471,348
Collection entity total	501,874	337,533
Establishment total	453,175	255,564
Tax Admin	518,943	509,985

This shows four different measures of operating revenue and operating expenses for the same enterprise. The first three sources are from surveys, the tax admin source represents the sum of revenue and expenses reported to the Canadian revenue agency for all legal entities within the enterprise. It is clear that by assembling the data in this manner an analyst can now begin the task of trying to understand the data being received from these large enterprises and why discrepancies may exist.

### 2.2.3 Data Analysis

Two types of analysis exist. The first is the case-by-case type of analysis outlined above in the section on KPM enterprises. This is the in-depth type of analysis which results in records being corrected within the current processing stage. The second type of analysis focuses on the entire group of complex enterprises and attempts to find systematic problems within the process that need to be corrected and which will in turn lead to more coherent data. From our preliminary analysis of the 1997 UES data we found three main problems:

**Coverage** – Incoherence exists because the current structure on the business register is incomplete or the reporting arrangement did not cover all units. All things being equal, one would expect that the sum of revenue reported for the establishments within the enterprise would be greater than the revenue reported by the enterprise<sup>2</sup>. Therefore when the revenue for the enterprise is much larger than the revenue reported by the establishments, it may indicate a coverage problem due to missing business units. It might also indicate incomplete reporting by respondents possibly due to misunderstanding of which business units to include on the survey.

**Inter-Enterprise Activity** – Incoherence exists because the production units are including internal sales and expenses in their estimates of operating revenue and expenses that are netted out in the financial data.

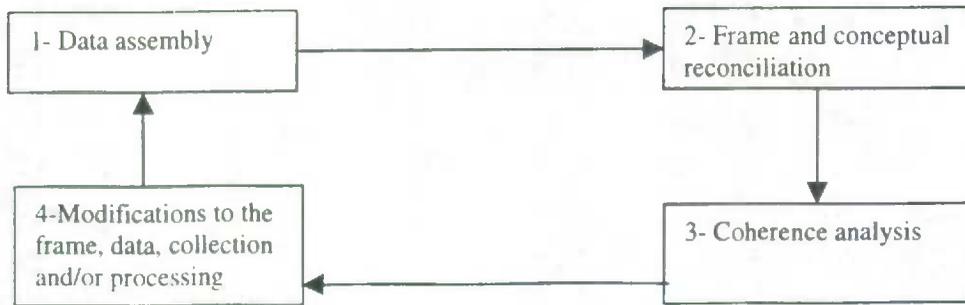
**Processing Systems** – Differences in the way data are edited, imputed, and allocated impacts the coherence of the data. Often even large enterprises will only provide partial responses to questionnaires and data need to be imputed. If the method is based upon donor imputation or nearest neighbor, there is nothing to ensure that the sum of the revenue (report and imputed) approximates the revenue reported by the enterprise.

## 3. Conclusion

While we have made great strides in measuring incoherence and understanding why it occurs, much work remains. Procedures need to be put in place which will ensure that the appropriate changes are made to the processing system to ensure a gradual progression to more coherent data. For this to happen, two things must occur. The first revolves around cooperation. Since the business survey program is organized by industry and many of the complex firms cross industry boundaries, analysts and processing systems must take this into account. At all stages of analysis and processing, cooperation must exist which leads to coherent data. Secondly, there must be accountability for overall coherence. As an agency, we are very accountable when it comes to a specific industry estimate, but it is the System of National Accounts, which becomes accountable for ensuring that the relationships between the various industries are coherent. This must be done earlier in the process and coherence needs to become part of the overall data quality assessment. Finally we must not rest on our laurels. The coherence process is cyclical (see diagram below) consisting of assembling the data, reconciling frame and conceptual inconsistencies, analysis, data correction (if required), frame updates, fine tuning the collection and/or processing system and then starting all over again. Once the data are coherent there is nothing to ensure that it will remain coherent.

---

<sup>2</sup> This is mainly because the enterprise revenue would include consolidation adjustments.



## 4. References

Bernier, R. and Mandeville, J "Coherence Analysis - Case study from the Key Provider Manager Program" – PIPES Technical Series Number 31, April 1999.

## Technical Series - Index

October 2, 2000

PIPES has a series of technical paper reprints primarily for internal purposes. A list of the reprints currently available is presented below. For copies, contact Bonnie Bercik at (613) 951-6790 or Diane Proulx at (613) 951-7192, fax number (613) 951-0411 or write to Statistics Canada, 13<sup>th</sup> Floor, Jean Talon Building, Tunney's Pasture, Ottawa, Ontario, K1A 0T6

1. Unified Enterprise Statistics Program – Project to Improve Provincial Economic Statistics – May 5, 1997 – PIPES Project Managers.
2. PIPES Evaluation Framework – September 15, 1997 – Philip Smith.
3. Report on the Unified Enterprise Survey & Reporting Arrangements Business Consultations – August 1997 – Guy Gellatly, Larry Murphy and Junior Smith.
4. Update on PIPES Progress: Notes for a Briefing for Federal and Provincial Finance Officials, Halifax, Nova Scotia, March 12, 1997 – Philip Smith.
5. An Overview of The Project to Improve Provincial Economic Statistics – November 1997 – George Beelen, Francine Hardy and Don Royce.
6. Using Databases to Design, Generate and Store Business Questionnaires at Statistics Canada – November 5, 1997 – Alana M. Boltwood.
7. The How and Why of Business Statistics – January 1999 – Elise Mennie. *(Not for external dissemination)*
8. An update on PIPES Fifteen Months into the Project – April 24, 1998 – Philip Smith.
9. Key Provider Manager (KPM) – 1997-98 Annual Report – May 1998 – Vicki Crompton.
10. A Framework for Planning Unified Enterprise Survey Data Collection – October 28, 1998 – Alana Boltwood.
11. Impact of the PIPES Funding on the Services Division Programme and Achievements in 1997-98 – April 1998 – Gordon Baldwin. *(Not for external dissemination)*
12. PIPES Organization and Decision-Making Structure – August 17, 1998 – Philip Smith. *(Not for external dissemination)*
13. The Central Goal of PIPES – November 17, 1997 – Philip Smith.
14. The Terminology and Framework of the Unified Enterprise Questionnaire – Revised March 1999 – Philip Smith.
15. Realizing and Measuring Quality Improvements in Provincial Economic Accounts – August 1998 – Philip Smith.
16. Annual Report 1997-98 – Ombudsman for Small Business Response Burden – July 1998, – Michael Issa. *(Not for external dissemination)*
17. Decision Making in PIPES – October 1, 1998 – Philip Smith.
18. Task Force on Electronic Data Reporting – April, 1998 – George Andrusiak, Monique Gaudreau, Laurie Hill, Anne Ladouceur, Denis Leblanc, Mario Ménard, Guy Parent, Joe Wilkinson, Doug Zinnicker.
19. PIPES Information Package – October 1998 – Philip Smith.
20. UES and the Non-Business Sectors – September 17, 1997 – Art Ridgeway.
21. CATS User Guide – April 1998 – Janet Howatson. *(Not for external dissemination)* Not available

22. Report on Collection and Data Capture Operation OID for UES 1997 – September 3, 1998 – Anne Ladouceur. *(Not for external dissemination)*
23. SDD Contribution to PIPES 1998-1999 – September 1998 – Shirley Dolan.
24. The Harmonized Sales Tax Revenue Allocation Formula – August 1998 – Karen Hall. *(Not for external dissemination)*
25. Task Group on Data Acquisition Strategy Report – July 22, 1998 – François Maranda and Don Royce.
26. Roles and Responsibilities in the Unified Enterprise Statistics Program – December 15, 1998 – George Andrusiak, Richard Barnabé, Albert Meguerditchian, Ray Ryan and Philip Smith. *(Not for external dissemination)*
27. Paper on the Project to Improve Provincial Economic Statistics from the Joint IASS/IAOS Conference – July 22, 1998 – Don Royce.
28. Respondent Relations Task Force – March 5, 1999 – Wayne Smith.
29. Response Analysis Follow-up Survey – March 1999 – Kristen Underwood.
30. Data Sharing Information Package – March 1999 – John Crysdale. *(Not for external dissemination)*
31. Coherence Analysis – Case Study from the Key Provider Manager Program – April 23, 1999 – Rachel Bernier and Julie Mandeville.
32. Evaluation of Collection Support Material used during the 1997 Unified Enterprise Survey – November 16, 1998 – Yvele Paquette.
33. Waiver Information Package – May 1999 – John Crysdale. *(Not for external dissemination)*
34. The PIPES Plan for 1999-00 – June 14, 1999 – Philip Smith. *(Not for external dissemination)*
35. BTS + Forum Post-conference Actions – April 1999 – Cornwall Conference Participants.
36. Report of the Task Force on Sources of Business Information – March 1999 – Vicki Crompton and Mark Marcogliese.
37. Field 5 Task Force Report on Improving Generic Boards – August 1999 – Mel Jones.
38. Study of Business Survey Questionnaires – June 1999 – Jason Gilmore.
39. Complexity Scale for Business Questionnaires – June 1999 – Jason Gilmore.
40. Update on PIPES – September 1999 – Philip Smith.
41. Exclusion Thresholds & Sampling Practices for Business Surveys – Implementation Strategy – September 1999 – Implementation Strategy Team.
42. Use of Tax Data in the Production of Provincial Economic Statistics – October 1999 – Peter Bissett.
43. Data Quality Survey 1996 – March 1999 – Ed Bunko. *(Not for external dissemination)*
44. Estimates of Information Cost to Business Respondents, 1998 – September 16, 1999 – Linda Grant and Michael Issa. *(Not for external dissemination)*
45. Data Security Task Force – January 1999 – Mark Steski. *(Not for external dissemination)*
46. Key Provider Manager Program – 1998-99 Program Report – June 1999 – Monique Gaudreau.
47. Treatment of Head offices and other support units in the UES – October 1999 – Charles Delorme.

48. UES Walkthrough – December 6, 1999 – Philip Smith.
49. The UES Strategy to Combine Enterprise, Establishment and Legal-Entity Data – November 29, 1999 – Philip Smith.
50. The Data Quality of the 1997 Unified Enterprise Survey (UES) Pilot – January 10, 2000 – Stuart Pursey.
51. Options for the Survey of Household Spending – January 2000 – Cynthia Baumgarten, David Binder, Louis-Marc Ducharme, Alison Hale, Albert Meguerditchian, Mike Sheridan, M.P. Singh, Philip Smith, Maryanne Webber, Brian Williams, Karen Wilson. *(Not for external dissemination)* AWAITING APPROVAL BY MIKE SHERIDAN
52. PIPES Budget Targets for 2001-02 and Beyond – March 28, 2000 – Project Management Team. *(Not for external dissemination)*
53. Unincorporated Business Statistics from the Tax Estimates Program – March 8, 2000 – Philip Smith.
54. PIPES and the Use of Statistics for Administrative Purposes – April 12, 2000 – Philip Smith.
55. Information Manual for Interviewers on the 1999 Survey of the Construction Industry – February 2000 – Claude Grenier. *(Not for external dissemination)*
56. The Conceptual and Practical Evolution of the Unified Enterprise Survey Integrated Questionnaire – May 2000 – Mario A. Vella and Annette Laurent.
57. Statistics Canada's Broad Strategy for Business Statistics – May 25, 2000 – Philip Smith.
58. Calendarizing Business Survey Data – June 2000 – Peter S. Tarassoff.
59. Report of the Task Force on Disclosure Avoidance Practices for Business Data – October 2000 – Daphne Bennett, John Crysdale, Ziad Ghanem, Pat Grainger, Tony Labillois, Serge Lavallée, Jackie Leblanc, Robert Masse, Bruno Pépin, Randall Sheldrick, Leslie Shinder, Patricia Whitbridge and Elaine Wilson. *(Not for external dissemination)*
60. Coherence Analysis of Large Enterprise Data – June 2000 – Monique Gaudreau and Jim Tebrake.

## **Other documents**

PIPES Project Structure – Revised July 1999  
Confidentiality of Statistical Data – April 1995 – D. Binder and L. Desramaux  
The Business Register Hellerman Autocoder Project – November 1999 – Business Register Division  
(English version only)  
The Why of Business Surveys – March 2000

The Business Register Hellerman Autocoder Project – novembre 1999 – Division du registre des entreprises  
(version anglaise seulement)  
Le pourquoi des enquêtes-entreprises – mars 2000

44. Estimations des coûts d'information pour les entreprises répondantes, 1998 – le 16 septembre 1999 – Linda Grant et Michael Issa. (*Diffusion interne seulement*)
45. Groupe de travail de la sécurité des données – janvier 1999 – Mark Steski. (*Diffusion interne seulement*)
46. Programme des gestionnaires des répondants clés – Rapport de programme pour 1998-1999 – juin 1999 – Monique Gaudreau.
47. Traitement des sièges sociaux et autres unités de support dans l'EUE – le 28 octobre 1999 – Charles Delorme.
48. Tour d'horizon de l'EUE – le 6 décembre 1999 – Philip Smith.
49. Stratégie de l'EUE en vue de combiner les données au niveau de l'entreprise, de l'établissement et de l'entité légale – le 29 novembre 1999 – Philip Smith.
50. La qualité des données de l'Enquête-pilote unifiée de 1997 auprès des entreprises (EUE) – le 10 janvier 2000 - Stuart Pursey.
51. Options pour l'Enquête sur les dépenses des ménages – janvier 2000 – Cynthia Baumgarten, David Binder, Louis-Marc Ducharme, Alison Hale, Albert Meguerditchian, Mike Sheridan, M.P. Singh, Philip Smith, Maryanne Webber, Brian Williams, Karen Wilson. (*Diffusion interne seulement*) EN ATTENTE DE L'APPROBATION DE MIKE SHERIDAN
52. Budget projeté du PASEP pour les exercices 2001-2002 et ultérieurs – le 28 mars 2000 – Équipe de gestion de projet. (*Diffusion interne seulement*)
53. Statistiques sur les entreprises non constituées en société du Programme des estimations fiscales – le 8 mars 2000 – Philip Smith.
54. Le PASEP et l'utilisation des statistiques à des fins administratives – le 12 avril 2000 – Philip Smith.
55. Manuel d'information des intervieweurs – Enquête de 1999 sur l'industrie de la construction – février 2000 – Claude Grenier. (*Diffusion interne seulement*)
56. Évolution conceptuelle et pratique du questionnaire intégré de l'Enquête unifiée auprès des entreprises – mai 2000 – Mario A. Vella et Annette Laurent.
57. Stratégie générale de Statistique Canada relativement à la statistique des entreprises – le 25 mai 2000 – Philip Smith.
58. La calendrialisation des données des enquêtes-entreprises – juin 2000 – Peter S. Tarassoff.
59. Rapport du Groupe de travail sur les pratiques de contrôle de la divulgation des données sur les entreprises – octobre 2000 – Daphne Bennett, John Crysedale, Ziad Ghanem, Pat Grainger, Tony Labilloy, Serge Lavallée, Jackie Leblanc, Robert Masse, Bruno Pépin, Randall Sheldrick, Leslie Shinder, Patricia Whitbridge et Elaine Wilson. (*Diffusion interne seulement*)
60. Analyse de la cohérence des données sur les grandes entreprises – juin 2000 – Monique Gaudreau et Jim Tebrake.

## Autres documents

La structure du projet du PASEP – Révisé en juillet 1999

La confidentialité des données statistiques – avril 1995 – D. Binder et L. Desramaux

## Analyse de la cohérence des données sur les grandes entreprises

Les grandes entreprises ont un impact important sur l'économie canadienne et sont par conséquent susceptibles de figurer dans la plupart des enquêtes auprès des entreprises de Statistique Canada (SC). Des déclarations incomplètes ou inexactes par ces entreprises peuvent avoir des répercussions substantielles sur la qualité des estimations produites par le Bureau. Il est par conséquent important de s'assurer que les données d'enquête fournies pour les diverses unités de l'entreprise correspondent aux données de l'état des résultats consolidé et du bilan fournies pour l'entreprise dans son ensemble.

Depuis quelques années, Statistique Canada analyse la cohérence de ses statistiques sur les entreprises. Cette analyse a pris la forme d'études de cas au sujet de grandes entreprises, ou d'analyses des données sur les sièges sociaux et les unités auxiliaires, ainsi que des données concernant des petites entreprises complexes. Le présent document montre les faits saillants de ces analyses et énonce les principaux obstacles à la production de données cohérentes, ainsi que les mesures et les processus qu'un organisme statistique peut adopter pour produire des données plus cohérentes.

### Note de reconnaissance

*Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.*

Pour plus de renseignements sur ce document, veuillez communiquer avec Bonnie Bercik (613) 951-6790 ou Diane Proulx (613) 951-7192  
Télécopieur : (613) 951-0411

# Analyse de la cohérence des données sur les grandes entreprises

Cette communication a été présentée à la Deuxième Conférence internationale sur les enquêtes-établissements (CIEE-II) à Buffalo, New York, le 19 juin 2000

Monique Gaudreau, Jim Tebrake, Statistique Canada  
Statistique Canada, immeuble Principal, bureau 1105, Ottawa (Ontario) Canada K1A 0T6  
gaudmon@statcan.ca tebrajam@statcan.ca

**Mots-clés :** analyse de la cohérence, incohérence, entreprise complexe, Projet d'amélioration des statistiques économiques provinciales (PASEP), Enquête unifiée auprès des entreprises, gestionnaires des répondants clés.

## 1. Contexte

Au début de 1997, Statistique Canada lançait un vaste et important « Projet d'amélioration des statistiques économiques provinciales » (PASEP). De façon générale, l'objectif du PASEP consiste à augmenter de façon permanente et très substantielle le volume et la qualité des données statistiques provinciales et territoriales. Des améliorations seront apportées dans quatre domaines distincts de la qualité des données : couverture, dénombrement, uniformité et cohérence.

On parle de cohérence lorsque les données sur les ventes d'une entreprise (données relatives à l'offre) sont en équilibre avec les données sur les dépenses des consommateurs (données relatives à la demande). *En outre, les données recueillies auprès des établissements, de même que celles recueillies auprès de l'entreprise mère, doivent être uniformes au niveau interne, exhaustives, et exemptes de chevauchements.*

Afin de permettre la mise en œuvre du PASEP, Statistique Canada est à remanier l'ensemble de sa base de sondage pour les enquêtes auprès des entreprises. Avant 1997, Statistique Canada effectuait environ 200 enquêtes auprès des entreprises chaque année, au moyen de 700 questionnaires distincts. Ces enquêtes étaient confiées à différents gestionnaires de programme, regroupés en environ une dizaine de divisions spécialisées, chacune étant responsable d'un secteur particulier, comme l'agriculture, les transports ou la fabrication. Pour plusieurs enquêtes, les processus de production étaient intégrés verticalement, c'est-à-dire qu'ils étaient complètement pris en charge par les divisions spécialisées. Les processus de production comportaient diverses étapes, comme la préparation et l'envoi par la poste des questionnaires, la saisie des données, la pondération, l'agrégation et la totalisation des données, l'analyse des résultats et la préparation des produits d'information à diffuser. Pour certaines enquêtes, des étapes, comme la saisie des données ou l'envoi par la poste des questionnaires, étaient centralisées dans des divisions de « service » spécialisées. Cette façon de faire est généralement plus efficace et a aussi comme avantage qu'elle permet une plus grande cohérence entre les enquêtes ou les secteurs.

Statistique Canada encourage actuellement l'utilisation de concepts, de méthodes et de procédures uniformes pour ses enquêtes auprès des entreprises, grâce à divers mécanismes. Il s'agit notamment de cadres conceptuels normalisés (comme le Système de comptabilité nationale, SCN), de systèmes de classification normalisés (comme le Système de classification

des industries de l'Amérique du Nord), d'un registre commun des entreprises, d'un bassin commun de personnel pour les activités relatives à la méthodologie, à l'exploitation et au développement de systèmes, ainsi que de politiques d'enquête s'appliquant à tout le Bureau. Néanmoins, des manques d'uniformité et des chevauchements de contenu subsistent entre les enquêtes auprès des entreprises de Statistique Canada. Les données recueillies auprès des entreprises souffrent aussi d'un certain degré d'incohérence.

À partir de 1997, un nombre important des 100 enquêtes annuelles a fait l'objet d'une intégration échelonnée à un programme d'enquête principal et unique appelée Enquête unifiée auprès des entreprises (EUE). La nouvelle approche intégrée met l'accent sur les entreprises, ce qui fait que les données financières reçues du siège social d'une entreprise, par exemple, les bénéfices, correspondent aux données sur la production et sur les ventes reçues des divers établissements. Dans le cadre de l'EUE, on recueille davantage de détails selon la branche d'activité et la marchandise au niveau provincial, et on évite le chevauchement entre les divers questionnaires d'enquête.

Statistique Canada a créé le programme des gestionnaires des répondants clés (GRC), dans le cadre de son nouveau Projet d'amélioration des statistiques économiques provinciales (PASEP). Les objectifs de ce programme consistent à améliorer les relations de travail de Statistique Canada avec les entreprises importantes et complexes, la réponse, du point de vue de l'actualité, de l'intégralité et de la précision, ainsi que la qualité et l'uniformité des données déclarées. À cet égard, les gestionnaires des répondants clés doivent s'assurer que les données recueillies auprès des entreprises dont ils sont chargés peuvent être traitées comme un ensemble de données cohérentes et interreliées. La capacité d'analyse de la gamme complète des données pour des entreprises particulières entraînera une plus grande cohérence des données au niveau interne et servira de base pour donner suite à d'autres enjeux, comme des méthodes d'imputation de rechange.

## **2. Pourquoi l'analyse de la cohérence au niveau de l'entreprise est-elle importante?**

Pourquoi un organisme statistique devrait-il se préoccuper de la production de données cohérentes pour les entreprises sur une base individuelle? Après tout, son rôle principal ne consiste-t-il pas à produire des estimations de grande qualité au niveau de la province ou de la branche d'activité, en vue de leur utilisation dans le cadre du système de comptabilité nationale. La cohérence des données pour une entreprise ne devrait pas constituer une préoccupation majeure. Ou le devrait-elle? Dans la prochaine section, nous soulignons les cas où il est approprié d'entreprendre une analyse de la cohérence et nous traitons des avantages d'une telle analyse.

Au Canada, le paysage économique est constitué principalement d'un petit nombre de très grandes entreprises. Étant donné que l'activité des entreprises a tendance à être asymétrique dans la plupart des branches d'activité, certaines entreprises clés ont des répercussions importantes sur les estimations finales selon la branche d'activité. Dans les cas où des branches d'activité sont dominées par quelques grandes entreprises, une analyse de la cohérence est nécessaire. L'analyse de la cohérence au niveau de l'entreprise est nécessaire parce qu'elle comporte deux avantages majeurs.

Tout d'abord, elle permet de déceler la plus grave des erreurs, à savoir le double compte. Prenons le cas de l'entreprise A qui a reçu des questionnaires d'enquête au sujet de ses activités de fabrication et de commerce de gros, qui a déclaré des données pour les cinq dernières années, qui a été cohérente d'année en année, mais qui a déclaré les mêmes chiffres pour

l'ensemble de ses activités, tant dans le questionnaire de la fabrication que dans celui du commerce de gros, parce qu'elle ne pouvait fournir la ventilation demandée.

Si l'on ne procède pas à une forme ou une autre d'analyse de la cohérence, il n'existe aucune garantie que l'on pourra déceler le double compte. Cela aura pour effet que les estimations pour le commerce de gros comprendront des activités de fabrication et que les estimations pour la fabrication comprendront des activités du commerce de gros. Comme on le verra plus tard, le recours à une approche axée sur l'entreprise pour l'analyse permettra de déceler des erreurs de ce type.

Une telle analyse a aussi pour avantage qu'elle permet de déceler le sous-dénombrement. Il existe un défi commun pour toutes les enquêtes auprès des entreprises, à savoir assurer la mise à jour de la base de sondage. Une des façons de vérifier la couverture consiste à comparer le total au niveau de l'entreprise à la somme des parties. Si le total est plus important, cela peut vouloir dire qu'il y a sous-dénombrement. On peut bien illustrer ce point en procédant à une analogie avec le type de vérification qui est effectuée dans le cadre de la plupart des enquêtes. La plupart des enquêtes comporte une section sur les revenus et, dans la plupart des cas, on demande à la fois les revenus totaux (total de contrôle) et les divers postes qui les constituent. Si le total et la somme des composantes individuelles ne sont pas égaux, on procède à une imputation. Dans le cadre de l'analyse de la cohérence, les revenus totaux correspondent au total pour l'entreprise, et les diverses parties correspondent aux différents établissements de l'entreprise. Si l'on ne compare pas le total pour l'entreprise à la somme des parties, il est impossible d'assurer la couverture complète de l'entreprise.

On pourrait prétendre que les deux avantages mentionnés ci-dessus ont tendance à s'annuler mutuellement, et que l'analyse de la cohérence au niveau de l'entreprise n'est donc pas réellement importante. Toutefois, ces grandes entreprises sont tellement distinctes que cet effet d'annulation est peu probable.

## **2.1 Entreprises visées par le programme des GRC**

Les grandes entreprises ont un impact important sur l'économie canadienne et sont par conséquent susceptibles d'être incluses dans les diverses enquêtes de Statistique Canada. Il importe par conséquent de s'assurer que des données cohérentes sont recueillies, tant au niveau de l'unité financière que de l'unité de production, et que ces données sont prises en compte avec précision dans les données sur l'économie canadienne.

Une étude de cas a été entreprise au sujet de deux sociétés importantes visées par le programme des gestionnaires des répondants clés. L'analyse a consisté à comparer la somme des valeurs déclarées au niveau de l'établissement à la valeur déclarée au niveau de l'entreprise, en tenant compte des ventes et des transferts internes et externes. Parmi les deux entreprises sélectionnées, une était entièrement intégrée à l'EUE, ce qui signifie que tous ses établissements faisaient partie de l'échantillon de l'EUE, tandis que l'autre n'était intégrée que partiellement, ce qui signifie que certains établissements ne faisaient pas partie de l'échantillon de l'EUE. Sept variables clés ont été utilisées pour l'analyse :

1. dépenses au titre des biens et services (avant l'EUE);  
total des dépenses d'exploitation (après l'EUE);
2. revenu de la vente de biens et de services (avant l'EUE);  
total des revenus d'exploitation (après l'EUE);
3. bénéfices bruts d'exploitation;
4. nombre d'employés;
5. salaires et traitements;
6. dépenses en immobilisations;
7. stocks d'ouverture et de fermeture.

Les représentants de l'entreprise totalement intégrée à l'EUE ont procédé à l'analyse eux-mêmes et ont corrigé toutes les incohérences décelées. Il leur a fallu 60 heures pour rapprocher les données qui nous avaient déjà été fournies.

L'analyse des données de l'entreprise non intégrée à l'EUE a consisté à comparer la somme des valeurs déclarées au niveau de l'établissement aux valeurs déclarées au niveau de l'entreprise, en tenant compte des transferts et des ventes internes et externes. Les huit variables énumérées ci-dessus ont été analysées. Les écarts allaient de 1 % pour le total des revenus d'exploitation à 44 % pour les stocks, entre les données recueillies auprès de l'entreprise et celles recueillies auprès des établissements. Parmi les raisons qui ont donné lieu aux incohérences figurent des différences entre les concepts et les définitions utilisés par les secteurs d'enquête (avant l'EUE), des erreurs de couverture et de déclaration, des erreurs de traitement et des transferts.

Dans les deux cas, il ne s'agissait pas de déterminer si les données pouvaient être rapprochées ou non, puisqu'elles le pouvaient. Il s'agissait plutôt de savoir quelle somme de travail était nécessaire pour rapprocher les données, et quelles procédures on pouvait utiliser pour s'assurer que des incohérences ne se reproduisent plus à l'avenir.

L'étude a aussi donné lieu à des recommandations pour un cadre d'analyse de la cohérence relativement au programme des GRC<sup>1</sup>. Les recommandations peuvent être résumées comme suit :

1. on pourrait élaborer une approche systématique pour recueillir et présenter les données;
2. le calcul de la valeur ajoutée devrait être automatisé;
3. les données administratives devraient faire partie de l'analyse de la cohérence;
4. des seuils d'acceptation selon la variable devraient être définis.

## 2.2 Au-delà de l'étude de cas

L'analyse de la cohérence est un nouveau concept. Pour de nombreux aspects, les analystes ont pour la plupart procédé à une forme ou à une autre d'analyse de la cohérence par le passé. Il arrive souvent qu'un analyste responsable d'un répondant important dans le cadre d'une enquête s'informe des activités du même répondant dans le cadre d'une autre enquête. Il existe toutefois une lacune, dans nombre de cas à tout le moins, à savoir que l'on a souvent adopté une approche non structurée pour procéder à ce type d'analyse. On a rarement pris des mesures officielles ou concrètes à cette fin. La présente section du document énonce en détail les étapes nécessaires pour aller au-delà des études de cas et intégrer la cohérence au programme des enquêtes annuelles auprès des entreprises. De façon plus particulière, nous avons déterminé trois processus nécessaires pour l'analyse de la cohérence, à savoir :

- le rapprochement;
- le regroupement des données;
- l'analyse des données.

### 2.2.1 Rapprochement

L'analyse de la cohérence consiste à regrouper les données. Même si en principe cela peut sembler une tâche uniquement technique, il ne s'agit en fait que l'un des aspects du travail. Avant que les données puissent être regroupées, trois types de rapprochements doivent être effectués.

- Rapprochement des bases de sondage

L'un des principaux obstacles à l'analyse de la cohérence vient du fait que souvent, la même base de sondage n'est pas utilisée pour toutes les enquêtes. Dans ces cas, un exercice

<sup>1</sup> Voir le document n° 31 de la série des documents techniques du PASEP - Analyse de cohérence - Étude de cas du programme des gestionnaires des répondants clés.

manuel de couplage des données doit alors être entrepris. Il arrive souvent que cet exercice de rapprochement soit tellement long et fastidieux qu'il n'est pas effectué ou qu'il n'est pas effectué suffisamment rapidement pour être utile.

➤ **Rapprochement conceptuel**

Des problèmes surviennent aussi lorsque dans le cadre des diverses enquêtes, on pose les questions de façon différente, ou on tente de mesurer les choses de façon légèrement différente. Cela a pour résultat que l'on compare des pommes et des oranges, et que l'on doit faire certaines hypothèses pour comparer les données de façon précise.

➤ **Rapprochement de traitement**

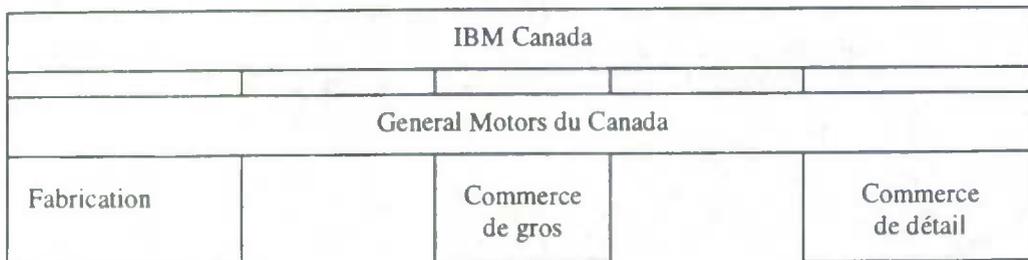
S'il réussit à coupler les bases de sondage et à résoudre les écarts conceptuels, l'analyste doit encore tenir compte des différences quant à la façon dont les données ont été traitées. La vérification et l'imputation ont-elles été fondées sur des enregistrements donneurs ou des enregistrements historiques? Si l'on a procédé à des vérifications historiques, a-t-on appliqué le même taux de croissance pour toutes les enquêtes auprès de la même entreprise? Les données ont-elles été réparties entre les provinces et, le cas échéant, a-t-on utilisé la même valeur d'imputation pour toutes les enquêtes.

Le programme de l'Enquête unifiée auprès des entreprises (EUE) à Statistique Canada a permis de supprimer le problème de rapprochement lié à l'analyse de la cohérence. Chaque enquête repose sur une base de sondage centrale, et un service centralisé permet de veiller à ce que les questions posées dans le cadre de chaque enquête soient aussi similaires que possible du point de vue conceptuel. Le traitement des données a été centralisé pour s'assurer que les mêmes méthodes sont utilisées pour les étapes de la saisie, de la vérification, de l'imputation et de la répartition. C'est cela plus que toute autre chose qui a rendu possible l'analyse de la cohérence.

**2.2.2 Regroupement des données**

Même si les changements mentionnés ci-dessus ont contribué de façon substantielle à rendre possible l'analyse de la cohérence, il est encore nécessaire de regrouper les données pour que l'analyse puisse être effectuée. Dans la majeure partie des cas, les données sont encore regroupées selon la branche d'activité. Il s'agit d'un choix logique, étant donné que l'objectif final consiste à produire des estimations fiables de la valeur ajoutée selon la branche d'activité. Par conséquent, les bases de données et les outils utilisés pour extraire les données sont structurés sur cette base. Toutefois, pour assurer la cohérence, les données doivent aussi être présentées dans une perspective axée sur l'entreprise.

Prenons l'illustration ci-dessous. Plutôt que de regrouper les établissements selon la branche d'activité (p. ex. groupe d'établissements du commerce de gros, groupe d'établissements de la fabrication), nous devons produire des bases de données qui regroupent les établissements à l'intérieur de leur structure d'organisation globale. Nous devons commencer par analyser l'enquête annuelle auprès d'IBM ou l'enquête annuelle auprès de General Motors. Puis, nous devons traiter les données, de la même façon que nous traiterions un groupe d'établissements au sein de la même branche d'activité. Cette façon de faire contribue pour une large part à la production de données cohérentes.



À l'heure actuelle, à Statistique Canada, nous disposons de nombreuses sources d'information différentes pour une entreprise particulière. Dans le cadre du projet de l'EUE, nous constituons une base de données axée sur l'entreprise, laquelle rend possible ce type d'analyse pour les différentes entreprises. Quelle que soit l'entreprise, nous disposons de la totalité ou d'une partie des renseignements suivants :

1. Base de sondage de l'entreprise (Registre central des entreprises)
2. Données sur l'entreprise tirées des questionnaires d'enquête axés sur l'entreprise
3. Base de sondage des établissements (Registre central des entreprises)
4. Données sur les établissements tirés des questionnaires d'enquête axés sur les établissements de l'EUE
5. Données financières sur les entités légales (déclarations de revenus des sociétés)
6. Salaires et traitements (données administratives sur les retenues à la source)

Même si un analyste doit pleinement comprendre ce que chacune de ces sources d'information tente de couvrir, il doit les comparer les unes aux autres afin d'obtenir un aperçu global des renseignements qu'une entreprise particulière fournit à Statistique Canada. La première portion de l'analyse de la cohérence consiste alors simplement à recueillir des données pour que l'analyste puisse disposer d'un tableau global. Voici un exemple d'un tel tableau.

#### Entreprise A

Source	Revenus d'exploitation (en milliers de \$)	Dépenses d'exploitation (en milliers de \$)
Total pour l'entreprise	475 499	471 348
Total pour l'entité de collecte	501 874	337 533
Total pour l'établissement	453 175	255 564
Données fiscales	518 943	509 985

On dispose de quatre mesures différentes des revenus et des dépenses d'exploitation pour la même entreprise. Les trois premières sources sont des enquêtes, les données fiscales représentant la somme des revenus et des dépenses déclarés à l'agence des revenus du Canada pour toutes les entités légales de l'entreprise. Il est clair qu'en regroupant les données de cette façon, un analyste peut entreprendre l'analyse des données qui sont reçues des grandes entreprises, et tenter de déterminer les raisons des écarts.

### 2.2.3 Analyse des données

Il existe deux types d'analyses. La première est l'analyse cas par cas dont il est question ci-dessus dans la section sur les entreprises visées par le programme des GRC. Il s'agit d'une analyse approfondie qui donne lieu à la correction des enregistrements à l'intérieur de l'étape de traitement en cours. Le deuxième type d'analyse met l'accent sur le groupe complet des entreprises complexes et vise à trouver les problèmes systématiques qui doivent être corrigés à l'intérieur du processus, en vue de la production de données plus cohérentes. À partir de notre analyse préliminaire des données de l'EUE pour 1997, nous avons déterminé trois problèmes majeurs.

**Couverture** – L'incohérence découle de la structure incomplète du registre des entreprises ou encore du fait que les modalités de déclaration n'englobent pas toutes les unités. Toutes choses étant égales par ailleurs, on pourrait s'attendre à ce que la somme des revenus déclarés pour les

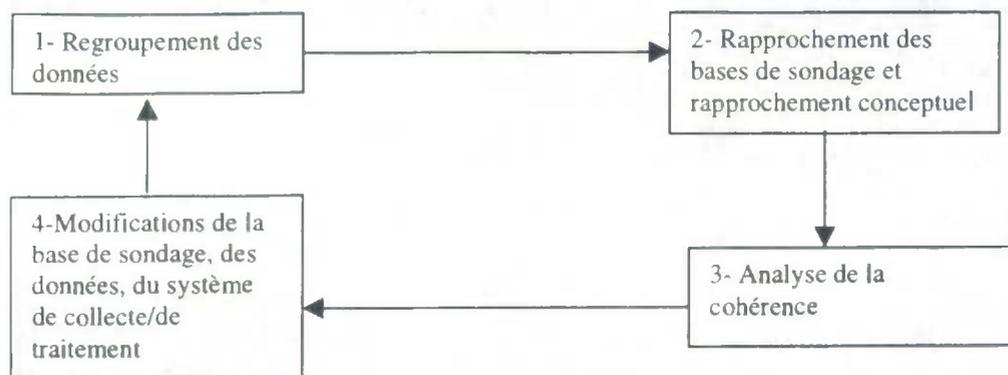
établissements de l'entreprise soit supérieure aux revenus déclarés par l'entreprise<sup>2</sup>. Par conséquent, lorsque les revenus de l'entreprise sont beaucoup plus importants que les revenus déclarés par les établissements, on peut avoir affaire à un problème de couverture découlant de l'absence de certaines unités de l'entreprise. On peut aussi avoir affaire à une déclaration incomplète par les répondants, laquelle peut être due à une mauvaise compréhension des unités à inclure dans l'enquête.

**Activité inter-entreprise** – L'incohérence découle du fait que les unités de production incluent les ventes et les dépenses internes dans leurs estimations des revenus et des dépenses d'exploitation qui sont déduites dans les données financières.

**Systèmes de traitement** – Les différences quant à la façon dont les données sont vérifiées, imputées et réparties a des répercussions sur la cohérence de celles-ci. Il arrive souvent que même les grandes entreprises ne fournissent que des réponses partielles aux questionnaires et que des données doivent être imputées. Si l'on utilise l'imputation par donneur ou par la méthode du plus proche voisin, il n'existe aucune façon de s'assurer que la somme des revenus (déclarés et imputés) correspond approximativement aux revenus déclarés par l'entreprise.

### 3. Conclusion

Même si nous avons fait de grands progrès quant à la mesure de l'incohérence et à la compréhension des raisons qui la justifient, il reste beaucoup à faire. Des procédures doivent être mises en place afin de s'assurer que des changements appropriés sont apportés au système de traitement pour permettre une progression graduelle vers des données plus cohérentes. Pour que cela se produise, deux éléments sont nécessaires, dont le premier est la collaboration. Étant donné que le programme des enquêtes auprès des entreprises est organisé selon la branche d'activité, et que nombre des entreprises complexes chevauchent plusieurs branches d'activité, les analystes et les systèmes de traitement doivent tenir compte de cet élément. À toutes les étapes de l'analyse et du traitement, il faut faire preuve de collaboration pour obtenir des données cohérentes. En deuxième lieu, quelqu'un doit assumer la responsabilité de la cohérence globale. En tant qu'organisme, nous sommes très responsables pour ce qui a trait aux estimations propres à une branche d'activité, mais ce sont les responsables du Système de comptabilité nationale qui doivent veiller à ce que les rapports entre les diverses branches d'activité sont cohérents. Cela doit être fait plus tôt dans le processus, et la cohérence doit faire partie de l'évaluation globale de la qualité des données. Enfin, nous ne devons pas nous reposer sur nos lauriers. Le processus de cohérence est cyclique (voir le diagramme ci-dessous) et consiste à regrouper les données, à rapprocher les bases de sondage et à résoudre les incohérences conceptuelles, à procéder à l'analyse, à corriger les données (au besoin), à raffiner le système de collecte ou de traitement, puis à reprendre le tout depuis le début. Une fois que les données sont cohérentes, il n'existe aucun mécanisme pour s'assurer qu'elles le demeureront.



<sup>2</sup> Cela vient principalement du fait que les revenus de l'entreprise incluraient les redressements de consolidation.

## 4. Bibliographie

Bernier, R. et Mandeville, J. « Analyse de cohérence - Étude de cas du programme des gestionnaires des répondants clés » – Document n° 31 de la série des documents techniques du PASEP, avril 1999.

Dans le cadre du PASEP, on a réimprimé une série de documents techniques, principalement pour usage interne. Voici la liste des réimpressions disponibles. Pour obtenir des copies communiquez avec Bonnie Bercik au (613) 951-6790 ou Diane Proulx au (613) 951-7192, numéro de télécopieur (613) 951-0411 ou écrivez à Statistique Canada, 13<sup>e</sup> étage, Immeuble Jean Talon, Parc Tunney, Ottawa, Ontario, K1A 0T6

1. Programme unifié des statistiques sur les entreprises – Projet d'amélioration des statistiques économiques provinciales – le 5 mai 1997 – Programme de Gestionnaire du PASEP.
2. Cadre d'évaluation du PASEP – le 15 septembre 1997 – Philip Smith.
3. Rapport de l'Enquête unifiée sur les entreprises et les modalités de déclaration – Consultations auprès des entreprises – août 1997 – Guy Gellatly, Larry Murphy et Junior Smith.
4. Bilan de l'évolution du PASEP : Notes d'une séance d'information à l'intention des représentants fédéral et provinciaux des finances, qui a eu lieu à Halifax (Nouvelle-Écosse), le 12 mars 1997 – Philip Smith.
5. Aperçu du Projet d'amélioration des statistiques économiques provinciales – novembre 1997 – George Beelen, Francine Hardy et Don Royce.
6. Des bases de données pour la conception, la génération et le stockage des questionnaires-entreprises à Statistique Canada – le 5 novembre 1997 – Alana M. Boltwood.
7. La statistique des entreprises : sa raison d'être – janvier 1999 – Elise Mennie. (*Diffusion interne seulement*)
8. Bilan du PASEP 15 mois après son lancement – le 24 avril 1998 – Philip Smith.
9. Programme des gestionnaires des répondants clés (GRC) – Rapport annuel pour 1997-1998 – mai 1998 – Vicki Crompton.
10. Un cadre de planification de la collecte des données de l'Enquête unifiée sur les entreprises – le 28 octobre 1998 – Alana Boltwood.
11. Répercussions du financement du PASEP sur le programme et les réalisations de la Division des services en 1997-1998 – avril 1998 – Gordon Baldwin. (*Diffusion interne seulement*)
12. L'organisation et la structure décisionnelle du PASEP – le 17 août 1998 – Philip Smith. (*Diffusion interne seulement*)
13. Les buts principaux du PASEP – le 17 novembre 1997 – Philip Smith.
14. Terminologie et cadre de référence du questionnaire de l'Enquête unifiée sur les entreprises – Révisé en mars 1999 – Philip Smith.
15. Amélioration de la qualité des statistiques économiques provinciales et mesure des changements apportés – août 1998 – Philip Smith.
16. Rapport annuel 1997-1998 – Médiateur – Fardeau de réponse de la petite entreprise – juillet 1998 – Michael Issa. (*Diffusion interne seulement*)
17. Le processus décisionnel du PASEP le 1<sup>er</sup> octobre 1998 – Philip Smith.
18. Groupe de travail sur la déclaration électronique des données (DED) – avril 1998 – George Andrusiak, Monique Gaudreau, Laurie Hill, Anne Ladouceur, Denis Leblanc, Mario Ménard, Guy Parent, Joe Wilkinson, Doug Zinnicker.
19. Trousse d'information sur le PASEP – octobre 1998 – Philip Smith.
20. L'EUE et les secteurs non commerciaux – le 17 septembre 1997 – Art Ridgeway.

21. Le guide de l'utilisateur du SASC – avril 1998 – Janet Howatson. (*Diffusion interne seulement*) Non disponible
22. Compte rendu de la collecte et la saisie de données DOI pour l'EUE de 1997 – le 3 septembre 1998 – Anne Ladouceur. (*Diffusion interne seulement*)
23. Contribution prévue de la DDS au PASEP, 1998-1999 – septembre 1998 – Shirley Dolan.
24. La formule de répartition des recettes de la taxe de vente harmonisée – août 1998 – Karen Hall. (*Diffusion interne seulement*)
25. Groupe de travail sur l'acquisition des données auprès des entreprises – le 22 juillet 1998 – François Maranda et Don Royce.
26. Rôles et responsabilités dans le cadre du Programme unifié des statistiques sur les entreprises – le 15 décembre 1998 – George Andrusiak, Richard Barnabé, Albert Meguerditchian, Ray Ryan et Philip Smith. (*Diffusion interne seulement*)
27. Document sur le Projet d'amélioration des statistiques économiques Provinciales de la conférence mixte de l'AISE/AISO – le 22 juillet 1998 – Don Royce.
28. Groupe de travail sur les relations avec les répondants – le 5 mars 1999 – Wayne Smith.
29. Enquête de suivi et d'analyse des réponses – mars 1999 – Kristen Underwood.
30. Dossier d'information sur le partage des données – mars 1999 – John Crysdale. (*Diffusion interne seulement*)
31. Analyse de cohérence – Étude de cas du programme des Gestionnaires des répondants clés – le 23 avril 1999 – Rachel Bemier et Julie Mandeville.
32. Évaluation des documents de soutien de la collecte utilisés durant l'Enquête unifiée sur les entreprises de 1997 – le 16 novembre 1998 – Yvele Paquette.
33. Trousse d'information sur les renoncations – mai 1999 – John Crysdale. (*Diffusion interne seulement*)
34. Plan du PASEP pour 1999-2000 – le 14 juin 1999 – Philip Smith. (*Diffusion interne seulement*)
35. Forum SCE + Mesures de suivi de la conférence – avril 1999 – Participants de la conférence de Cornwall.
36. Rapport du groupe d'étude des sources d'information sur les entreprises – mars 1999 – Vicki Crompton et Mark Marcogliese.
37. Rapport du Groupe de travail du secteur 5 sur l'amélioration de la dotation générique – août 1999 – Mel Jones.
38. Étude des questionnaires des enquêtes auprès des entreprises – juin 1999 – Jason Gilmore.
39. Échelle de complexité des questionnaires sur les entreprises – juin 1999 – Jason Gilmore.
40. Rapport sur l'avancement du PASEP – septembre 1999 – Philip Smith.
41. Seuils d'exclusion et méthodes particulières d'échantillonnage pour les enquêtes-entreprises – Stratégie de mise en oeuvre – septembre 1999 – L'Équipe de la stratégie de la mise en oeuvre.
42. Utilisation des données fiscales pour la production des statistiques économiques provinciales – octobre 1999 – Peter Bissett.
43. Enquête sur la qualité des données de 1996 – mars 1999 – Ed Bunko. (*Diffusion interne seulement*)



Projet d'amélioration des statistiques  
économiques provinciales

## Analyse de la cohérence des données sur les grandes entreprises

Série technique

Numéro 60

Project to Improve Provincial Economic  
Statistics

## Coherence Analysis of Large Enterprise Data

Technical Series

Number 60



Internet : [www.statcan.ca](http://www.statcan.ca)  
Intranet : <http://pasep>



Statistique  
Canada

Statistics  
Canada

Canada

STATISTICS CANADA

1010312351

20

STATISTICS CANADA LIBRARY  
BIBLIOTHÈQUE STATISTIQUE CANADA



1010312351