

68N0003XPB

no. 65

c. 2

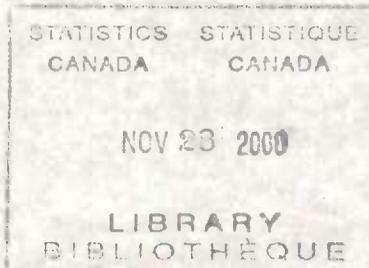
**Project to Improve Provincial Economic
Statistics**

A New Approach to Processing for the Unified Enterprise Statistics Program

**An Edit and Imputation
Example**

Technical Series

Number 65



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

Une nouvelle stratégie de traitement des données pour le programme unifié des statistiques auprès des entreprises

**Un exemple de vérification et
d'imputation**

Série technique

Numéro 65



Internet: www.statcan.ca
Intranet: <http://pipes>



Statistics
Canada

Statistique
Canada

Canada

A New Approach to Processing for the Unified Enterprise Statistics Program

An Edit and Imputation Example

The Unified Enterprise Survey (UES) is a recent initiative, designed to have all annual business surveys at Statistics Canada (STC) integrated and processed through a single processing system. The Edit & Imputation (E&I) sub-system is an important component of the UES. This paper addresses the challenges of managing a process that involves many diverse industries, the tight schedule and the systems development costs associated with such a large undertaking. The concept of integrating existing systems with new components and building generic meta-data driven systems is explored.

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

A New Approach to Processing for the Unified Enterprise Statistics Program

An Edit and Imputation Example

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

Colleen Martin and Jean-François Laroche, Statistics Canada
Colleen Martin, Statistics Canada Main Building, Ottawa, Ontario, K1A 0T6, Canada
martinc@statcan.ca

Key Words: Systems, Meta-data, Generic, Diverse

1. Introduction

The Unified Enterprise Survey (UES) was first implemented as a pilot for reference year 1997. A group of seven industries, not currently surveyed at Statistics Canada (STC), was selected for inclusion in this introductory year. It was the first year of a planned four year phased implementation that would result in about two thirds of all industries falling under the umbrella of a single annual business survey, supported by a single survey processing system.

From the point of view of industry coverage, frame and survey integration, response burden and various other content and subject matter aspects, it was anticipated that there would be many benefits as a consequence of this centralized approach. From the point of view of the computer systems development, it was seen as an opportunity to build one processing system to replace many independent processing systems and to subsequently realize benefits, particularly in the area of system maintenance.

The ambitious timetable established at the beginning of the project played a major role in determining the path taken through the computer systems design and implementation phases. For the edit and imputation (E&I) system, it meant that prototypes became 'one-off' production systems, re-use and re-cycle took on a new meaning and modularization became the mantra of the development team, all in an effort to meet the tight deadlines.

2. Challenges

At STC, most new survey development work is carried out in a project team environment. Historically, team members represent Business Survey Methods Division (BSMD), Systems Development Division (SDD) and one Subject Matter Area (SMA). For the UES, many SMAs are represented on the project team. Each area brings a culture different from its neighbour, a set of requirements it considers unique and a history of doing things its own way. It has been a

challenge to build a UES team culture wherein a common set of requirements and an acceptance of change can emerge.

Since the beginning of the project, the timetable set out for the UES has been challenging. For the pilot year, with 7 industries and 10 questionnaires involved, the E&I project team first convened approximately 14 months before the system went into production. Requirements gathering, systems development, testing and implementation of a prototype filled the gap. The system did the processing job it set out to do and provided an opportunity to test some long-term design theories. It was not itself a viable long-term solution. Although centralized, the system was still in many ways addressing the requirements of one industry at a time. It was clear that such methods were not feasible in the long run. We had neither the time nor the resources to prepare for year-2 with 10 new industries and 20 new questionnaires, in addition to the changes to the pilot questionnaires, using the same techniques.

The UES E&I system's greatest challenge had thus emerged. The number and complexity of the UES industry coverage would continue to grow for the next few years. The team therefore had to find a way to deal with new questionnaires and changes to old questionnaires efficiently, and with a minimum of intervention from systems developers. The major challenge for the design of the system was that of scalability, the ability to add new industries with a low marginal incremental cost. As the data content changed or questionnaires with completely new content were added, we could not afford to go back to systems developers to have programs modified or re-written. The team faced the task of developing a feasible long-term solution and again with little more than a year to put the first release into production.

3. Objectives and Constraints

Traditionally each area of subject matter expertise at STC has been responsible for building and maintaining its own processing systems. This has resulted in much duplication of effort, and duplication or near-duplication of actual system components. A major objective of the UES is to eliminate such duplication in all its systems, both manual and automated, through development of a common yet flexible overall approach that supports a diverse set of requirements for industries covered by the UES.

It is not the first time that we have considered this overall approach at STC. There is a series of generalized survey processing products, built at STC, and used by some, but by no means all, survey areas. The specific product that is relevant to the current topic is the Generalized Edit and Imputation System (GEIS). This system is composed of many useful imputation algorithms. Reluctance to use GEIS lies in the fact that it is difficult to execute on large data sets. It must be submitted one step at a time, for one imputation group at a time. In spite of this, it is an objective of the UES E&I project to take advantage, where possible, of existing GEIS modules.

Since the UES production will always occur in a very strict time frame, it is imperative that the automated system run efficiently and with as little human intervention as possible. No time can be lost. This implies that a method be found to accelerate execution such that pre-defined job networks can be submitted according to individual industry requirements.

An overriding influence, the greatest constraint, on the development of E&I systems is the time line, already established, for the UES project deliverables. Deadlines cannot be moved. If target dates are in jeopardy, functionality may be sacrificed in the event that resources are not available.

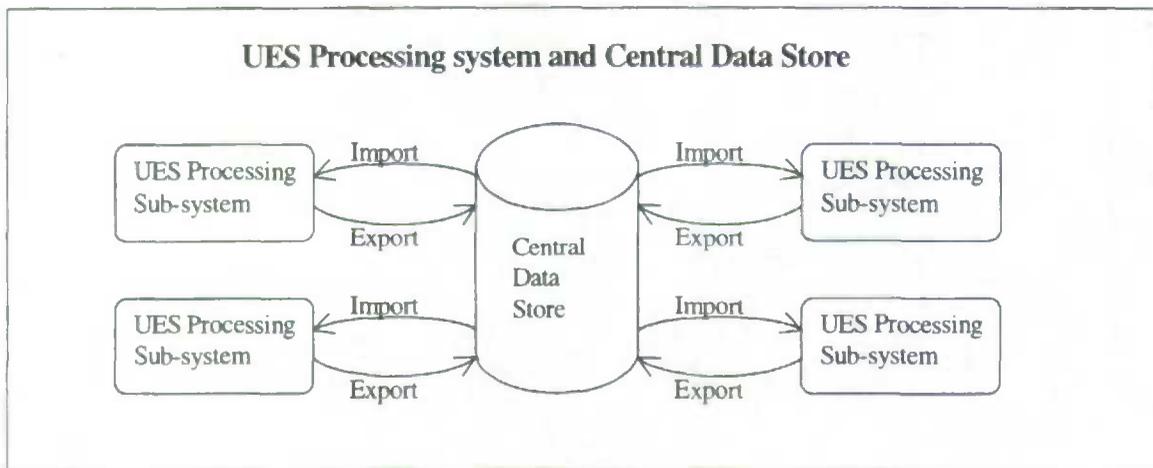
Finally, a most important objective, is one of continuous improvement. The systems and the methods that we put in place must support management's requirements for information about the processes and about the data transformations that take place through those processes. Only then

can appropriate decisions be taken regarding changes to the automated system and to the manual interventions, for future survey cycles.

4. The Approach

4.1 The Overall UES System Approach

The E&I sub-system is one of many components that make up the UES post-collection processing system. The first step in the development process was to develop an overall UES system design. The design that emerged was one that sees a Central Data Store (CDS), with its supporting import and export functions, surrounded by a series of satellite sub-systems, amongst them the E&I system. Each satellite system has its own data repository supported by the technology that best fits its processing requirements. Each execution of a processing component begins with an import of data from the CDS and ends with an export back to the CDS. The data exported from each sub-system are retained as a "snapshot" on the CDS, so that all transformations through the processing cycle for the data of an entire reference year can be followed.



4.2 The Edit and Imputation System Approach

Two main factors influenced the design of the edit and imputation system that is now in place for the UES... the requirement to have only one system process multiple diverse industries and the need to minimize development costs. The first release of the long-term UES E&I system had to support 17 industries, 30 questionnaires and in excess of 1,000 variables. For the second year, there will be approximately 25 industries and over 40 questionnaires. Further industries will be added in the third year, bringing the total number of questionnaires to more than 70. The system has to be able to process industries with diverse data content and to be flexible enough to accommodate addition of new industries or modification to the data content of existing industries from year to year with minimum involvement of IT professionals.

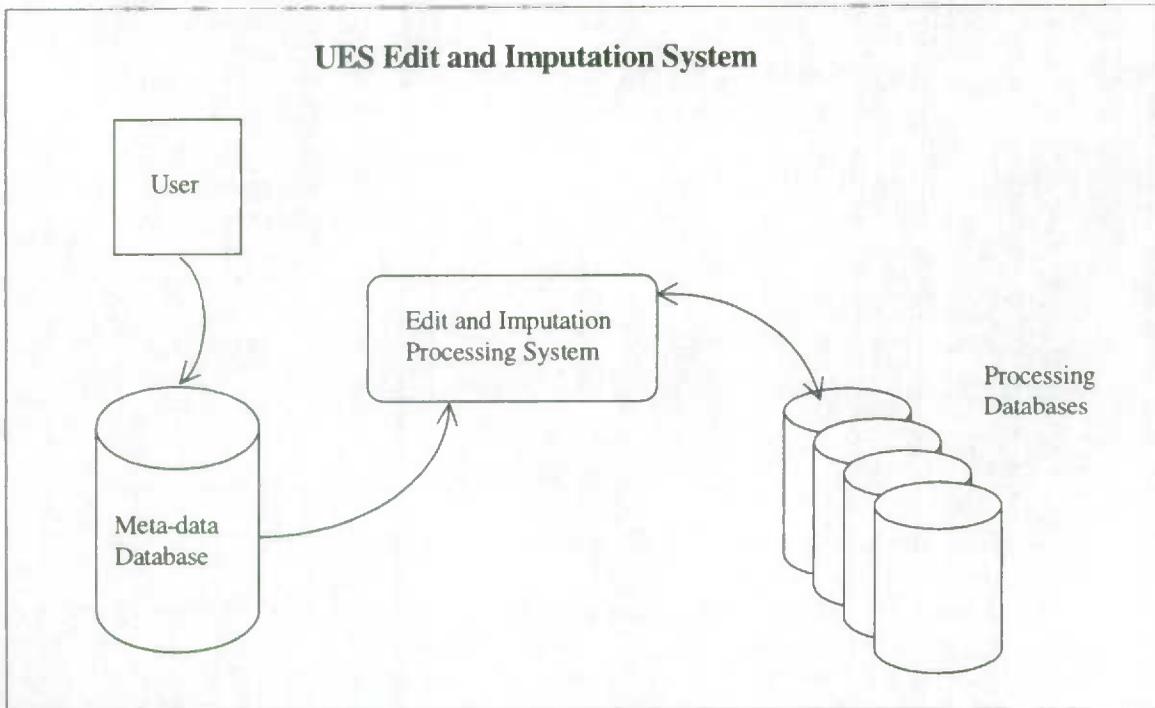
The approach taken to meet these requirements was to design a meta-data driven system, which aims to remove from the processing system all information specific to an industry. Essentially, with this design, the process applies a series of pre-defined algorithms to a set of abstract data elements. The knowledge about an industry is stored in a meta-data database to be accessed by the processing system when required.

4.2.1 Meta-data Database

The Meta-data Database satisfies three kinds of system requirements. Firstly, it defines the actual data content. It describes the data elements that are required to process each industry, in terms of current data, historical data, administrative data and so on. Secondly, it describes for each generic module or algorithm, the relationships between the data elements on which that algorithm must perform. Finally, it dictates which modules apply to an industry and in what order these modules must be applied.

4.2.2 Processing Databases

The Processing Database is the repository in which the data reside as they are being transformed by the E&I algorithms. For each industry, an instance of the E&I Processing Database is constructed and contains the survey data to be processed. This database is populated by importing the appropriate variables from the CDS. Following the import, the generic system modules are applied to the survey data stored in the Processing Database. The results are then exported back to the CDS and the database tables for the industry are optionally deleted. The meta-data that define the industry's data content support these import, export and table creation functions.



4.2.3 Edit and Imputation Processing System

The edit and imputation processing system performs three major functions. It applies algorithms to verify and if necessary to transform the survey data. It supplies auxiliary information about the transformations. It controls the process flow according to the needs of the industry being processed.

The major element is the set of generic modules. Each of these modules is completely independent of the data content. Each module is associated with those meta-data that support its specific function. For example, the module that checks whether the sum of reported details is equal to the reported total accesses those meta-data that identify groups of variables and their nested relationships, as well as those that identify the tolerance bounds. The module that identifies outliers to be excluded from the donor pool accesses those meta-data that identify the variables to be used in the calculation. In each case, the meta-data may differ from one industry

to another. There are in excess of 30 algorithms to meet the requirements of the first release of the system. It has been possible to use several modules of the GEIS system to satisfy certain of these requirements. The remaining algorithms have been built specifically for the UES. Since GEIS is also data content independent and meta-data driven, it fits extremely well in the context of the overall approach.

In order to manage and ultimately improve the conduct of the E&I process, the E&I manager needs to know which variables were transformed and by which algorithm. To this end, each survey data element that exists in the processing database has a corresponding status variable associated with it. When a data element is updated as a result of the application of a specific algorithm, its status variable is set to record the identification of the algorithm. At the end of the process, this information is available to be analyzed so that the E&I manager can assess the behaviour of the entire E&I function.

In order to satisfy the execution objectives of the system, the processing system includes an operational layer that controls which modules are used and in what order they are applied to a specific industry. The system has been described as "push-button". Even for GEIS modules, it is no longer necessary to stop at each processing step before continuing to the next.

4.2.4 The Process

To process an industry through the edit and imputation system:

- Users update the meta-data database with the information required by the edit and imputation system for the industry;
- Based on the meta-data database, an instance of the processing database is created for the industry to be processed and data are imported from the CDS;
- Based on the meta-data database, a set of pre-defined system modules is applied to the processing database of that industry and the data are exported back to the CDS.

5. Looking Back

With the first release of the long-term system solution now in production, there is a consensus that the meta-data approach is a correct one.

The diversity between industries can be managed through the meta-data database. The same system can easily process a large number of industries as long as these industries are correctly documented in the meta-data database. Modifications to the data content for an industry or addition of new industries can be managed by updating the meta-database without requiring the involvement of IT professionals.

The method has succeeded in eliminating E&I system duplication. Algorithms have been built only once and have been reused multiple times. Savings in initial development costs have been realized. The savings in maintenance costs for the future will be even more impressive.

GEIS modules have been incorporated successfully into an otherwise brand new system. The operational concerns previously associated with GEIS have been overcome, with the addition of an operational layer that predefines which modules to run and when to run them for each industry. This approach ensures that the actual processing timeframes can be met, since there is little or no human intervention between modules.

The decision to create and maintain a status variable for each survey variable has been well received. The status variables provide an opportunity for analysis at a level of detail that was not generally available in previous systems.

A few unplanned benefits have also emerged from following the extremely modular approach:

- It has allowed us, as deadlines loomed, to be selective about the order in which modules were finalized, since each module is totally independent of all others from the point of view of the physical implementation.
- Testing has been easier to manage.

The development project has not been entirely rosy, and the system does of course have limitations.

The generalization of the edit and imputation system modules is essentially content oriented. Variation in content can be easily managed by the system, but variation in processing requirements is another issue. If an industry requires a data transformation not currently available as one of the generic modules, a new generic module will have to be implemented to meet this new requirement. A large number of algorithms are already available in the system and it is expected that the number of new modules required in the future will be minimal.

Another important limitation of the system is related to the meta-data database. The meta-data database can be seen as a repository containing knowledge about the data for an industry and how to process this industry. In order to maintain it, a user must have a good knowledge of the data for the industry to be processed as well as a good knowledge of the processing system. The complexity of maintaining the meta-data database up-to-date should not be underestimated.

The meta-data approach has been more difficult to implement than a more traditional, subject-matter specific approach, from the point of view of both programmers and meta-data specifiers. There was a fairly steep learning curve for the programmers and a lot more meta-data maintenance than originally anticipated for the non-technical members of the team.

Establishing the UES E&I requirements, methods and priorities has been a bittersweet experience. In the pilot year, the Subject Matter Officers (SMOs) represented 7 industries that had not been previously surveyed. The SMOs for each industry were completely open to suggestions from other project participants. The team easily found consensus regarding requirements, methods and priorities. For the second year, which is now in production with the new system, and the third year, where requirements gathering has just been completed, the task has been a more difficult one. In each case, one or more industries have come to the UES with a survey history. For representatives of these industries, there is a different set of expectations than for those involved in the pilot. The quality and timeliness of the UES-produced survey estimate must match or exceed previous results. One way to ensure equality in this arena is to do things exactly as they were done before. In the meantime, a new UES way of doing things has begun to take hold amongst original project team members. There is often a struggle between the two points of view. Tensions are higher than in the pilot year. Consensus is more elusive.

6. Looking Forward

Maintaining meta-data has proven such a burden, that very high on the to-do list is the development of a user-friendly interface to support the effort. Since E&I does not exist in isolation, but rather as an integral part of a much larger UES system, the meta-data maintenance system will ultimately deal with the meta-data of all sub-systems. The eventual system must be intuitive and have checks and balances to disallow error of a systematic nature. The experience of defining the meta-data in a completely manual way has however been invaluable. We hope to do a better job of building such an interface having first done it the hard way.

We have already identified requirements for new algorithms for the second release of the system. Some of the requirements will again be met through GEIS modules, others will have to be built.

We plan to add even more flexibility in the selection of which modules to execute when. For the first release, we did a better job of this in connection with the GEIS modules than we did with the new UES modules.

We have built into the system the capacity to record information about the data transformations, through the status variables. We now need tools to help us analyze this information, so that the methods employed in the E&I system can be fine-tuned, or over-hauled, where necessary.

Technical Series - Index

November 16, 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, 13th 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 Whitridge and Elaine Wilson. (*Not for external dissemination*)
60. Coherence Analysis of Large Enterprise Data – June 2000 – Monique Gaudreau and Jim Tebrake.
61. An Overview of the 1998 GIFI Database – July 26, 2000 – Jennifer Sarah Jones.
62. Head Offices and Other Ancillary Units in the Unified Enterprise Survey – May 2000 – Malika Hamdad and Michel Tessier.
63. Exploring the Potential of EDA: Illustrated with examples from the Canadian Food Services Industry Annual Survey – June 2000 – Jim Tebrake, Daphne Bennett and Larry Murphy.
64. The Challenge of Collecting Quality Data from Large Enterprises: Lessons learned from the Key Provider Manager Approach – June 2000 – Monique Gaudreau and Janet Hughes.
65. A New Approach to Processing for the Unified Enterprise Statistics Program – An Edit and Imputation Example – June 2000 – Colleen Martin and Jean-François Laroche.

Other documents

PIPES Project Structure – Revised October 2000
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

64. Le défi que représente la collecte de données de qualité auprès des grandes entreprises : Enseignements tirés de la stratégie des gestionnaires des répondants clés – juin 2000 – Monique Gaudreau et Janet Hughes.
65. Une nouvelle stratégie de traitement des données pour le programme unifié des statistiques auprès des entreprises – Un exemple de vérification et d'imputation – juin 2000 – Colleen Martin et Jean-François Laroche.

Autres documents

La structure du projet du PASEP – Révisé en octobre 2000

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

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 Crysdale, Ziad Ghanem, Pat Grainger, Tony Labillois, Serge Lavallée, Jackie Leblanc, Robert Masse, Bruno Pépin, Randall Sheldrick, Leslie Shinder, Patricia Whitridge et Elaine Wilson. (*Diffusion interne seulement*)
60. Analyse de la cohérence des données sur les grandes entreprises – juin 2000 – Monique Gaudreau et Jim Tebrake.
61. Aperçu de la base de données de l'IGRF 1998 – le 26 juillet 2000 – Jennifer Sarah Jones.
62. Sièges sociaux et autres unités de soutien dans le cadre de l'Enquête unifiée auprès des entreprises – mai 2000 – Malika Hamdad et Michel Tessier.
63. Exploration des possibilités de l'EDA : Illustration à partir d'exemples de l'Enquête annuelle sur le secteur des services de restauration au Canada – juin 2000 – Jim Tebrake, Daphne Bennett et Larry Murphy.

Une nouvelle stratégie de traitement des données pour le programme unifié des statistiques auprès des entreprises

Un exemple de vérification et d'imputation

L'Enquête unifiée auprès des entreprises (EUE) est une initiative récente conçue pour intégrer et traiter toutes les enquêtes-entreprises annuelles réalisées par Statistique Canada (SC) au moyen d'un système de traitement unique. Le sous-système de vérification et d'imputation (V et I) est une composante importante du programme de l'EUE. Le présent article traite du défi que pose la gestion d'un processus qui couvre un grand nombre de branches d'activité forts diverses, d'un échéancier serré et des coûts de développement de systèmes qu'occasionne un entreprise de cette envergure. On y examine aussi l'intégration de systèmes existants et de nouvelles composantes, ainsi que la création d'un système générique axé sur des métadonnées.

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

Une nouvelle stratégie de traitement des données pour le programme unifié des statistiques auprès des entreprises

Un exemple de vérification et d'imputation

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

Colleen Martin et Jean-François Laroche, Statistique Canada
Colleen Martin, Immeuble principal, Statistique Canada, Ottawa (Ontario) K1A 0T6, Canada
martinc@statcan.ca

Mots clés : systèmes, métadonnées, générique, diverse

1. Introduction

L'Enquête unifiée auprès des entreprises (EUE) a été mise en œuvre pour la première fois à titre d'essai pilote pour l'année de référence 1997. On a sélectionné pour cet essai sept branches d'activité qui n'étaient couvertes à l'époque par aucune enquête de Statistique Canada (SC). L'enquête pilote devait représenter la première année d'une phase de mise en œuvre qui s'étendrait, en principe, sur quatre années au bout desquelles environ les deux tiers des branches d'activité seraient couvertes par une enquête unifiée annuelle auprès des entreprises appuyée par un système unique de traitement des opérations d'enquête.

En ce qui concerne la couverture des branches d'activité, l'intégration des bases de sondage et des enquêtes, le fardeau de réponse et divers autres aspects du contenu spécialisés, on s'attendait à ce que le projet offre de nombreux avantages en raison de la stratégie de centralisation. Du point de vue de la mise au point des systèmes informatiques, le projet était considéré comme une occasion de mettre en place un système de traitement unique pour remplacer de nombreux systèmes de traitement indépendants, stratégie qui devrait offrir des avantages, particulièrement dans le domaine de la maintenance des systèmes.

L'échéancier ambitieux établi au début du projet a été un facteur déterminant du cheminement suivi lors de la conception des systèmes informatiques et des phases de mise en œuvre. Ainsi, en ce qui concerne le système de vérification et d'imputation (V et I), les prototypes sont devenus des systèmes de production « uniques », les termes « réutilisation » et « recyclage » ont pris une nouvelle signification et la modularisation est devenue le leitmotiv de l'équipe de développement en vue de respecter les échéances serrées.

2. Défis

À SC, la plupart des travaux d'élaboration de nouvelles enquêtes sont exécutés en équipe. Habituellement, celle-ci compte des représentants de la Division des méthodes d'enquête-entreprise (DMEE), de la Division du développement des systèmes (DDS) et d'une division spécialisée. Dans le cas de l'EUE, un grand nombre de divisions spécialisée sont représentées au sein de l'équipe. Or, chacune de ces divisions apporte une culture qui lui est propre, un ensemble d'exigences qu'elle considère comme uniques et des pratiques particulières établies de longue date. Le défi a donc consisté à créer une culture propre à l'équipe de l'EUE de sorte que ses membres adoptent un ensemble commun d'exigences et acceptent les changements.

Dès le début du projet, l'échéancier a été exigeant. Pour l'essai pilote, qui couvrait sept branches d'activité et comptait 10 questionnaires, l'équipe de projet de la V et I s'est réunie pour la première fois 14 mois environ avant la mise en exploitation du système. La définition des exigences, le développement des systèmes, la mise à l'essai et la mise en œuvre d'un prototype ont eu lieu dans l'intervalle. Le système a produit les résultats de traitement attendus et donné l'occasion de mettre à l'essai certaines théories de conception à long terme. Le système ne représentait pas, en soi, une solution viable à long échéance. Quoique centralisé, de bien des façons, il répondait encore aux exigences d'une seule branche d'activité à la fois. Or, il était manifeste que ce genre de méthode ne pourrait être applicable à longue échéance. Nous ne disposions ni du temps ni des ressources nécessaires pour assurer les préparatifs de la deuxième année, durant laquelle on étudierait 10 nouvelles branches d'activité et on se servirait de 20 nouveaux questionnaires, sans compter les modifications apportées au questionnaire de l'essai pilote, si l'on continuait d'utiliser les mêmes méthodes.

Le défi le plus important que devrait relever l'équipe chargée du système de V et I de l'EUE venait donc de se poser. Le nombre et la complexité des branches d'activité incluses dans le champ d'observation de l'EUE continueraient d'augmenter d'année en année. Par conséquent, l'équipe devait trouver un moyen efficace de répondre à l'ajout de nouveaux questionnaires et à la modification des anciens en réduisant au minimum l'intervention des développeurs de systèmes. La plus grande difficulté que posait la conception du système tenait à la création d'une architecture flexible permettant d'ajouter de nouvelles branches d'activité moyennant un faible coût. Nous ne pouvions nous permettre de demander aux développeurs de systèmes de modifier ou de réécrire les programmes à mesure qu'évoluerait le contenu des données ou que seraient ajoutés des questionnaires entièrement nouveaux. L'équipe a donc dû mettre au point une solution viable à long terme, tout en disposant d'à peine plus d'un an pour mettre la première version en exploitation.

3. Objectifs et contraintes

Traditionnellement, chaque division spécialisée de SC était chargée de créer et de tenir à jour ses propres systèmes de traitement. Cette stratégie a donné lieu à un chevauchement important des efforts et à la répétition ou à la quasi répétition des composantes des systèmes. L'un des objectifs importants de l'EUE est d'éliminer ce genre de répétition dans tous les systèmes pertinents, manuels ou automatisés, grâce à l'élaboration d'une stratégie globale commune, quoique souple, qui répond à un ensemble varié d'exigences pour les branches d'activité couvertes par l'enquête.

Ce n'est pas la première fois que l'on envisage d'adopter ce genre de stratégie globale à SC. Le Bureau a conçu une série de produits généralisés de traitement des données d'enquête qui sont utilisés par certaines divisions spécialisées. Le produit particulier qui nous intéresse ici est le Système généralisé de vérification et d'imputation (SGVI). Ce système comprend de nombreux algorithmes d'imputation. L'hésitation à utiliser le SGVI tient au fait que son exécution est difficile sur de grands ensembles de données. Elle doit se faire une étape à la fois, pour un groupe d'imputations à la fois. Malgré cet inconvénient, l'équipe de projet de la vérification et de l'imputation de l'EUE est déterminée à tirer parti, dans la mesure du possible, des modules existants du SGVI.

Puisque l'échéancier de la réalisation de l'EUE sera toujours très stricte, il est impérieux que le système automatisé fonctionne bien et demande aussi peu d'interventions humaines que possible. Il ne faut absolument pas perdre de temps. Autrement dit, il faut trouver un moyen d'accélérer l'exécution, de sorte que les ensembles prédéfinis de tâches puissent être exécutés conformément aux exigences particulières des branches d'activité.

La plus grande contrainte, qui surpasse toutes les autres, imposée à l'équipe chargée du développement des systèmes de V et I est l'échéancier déjà établi pour la livraison des produits de l'EUE. Les échéances ne pouvant être reportées, si elles risquent de ne pas être respectées, il faut parfois sacrifier des fonctionnalités si des ressources supplémentaires ne sont pas disponibles.

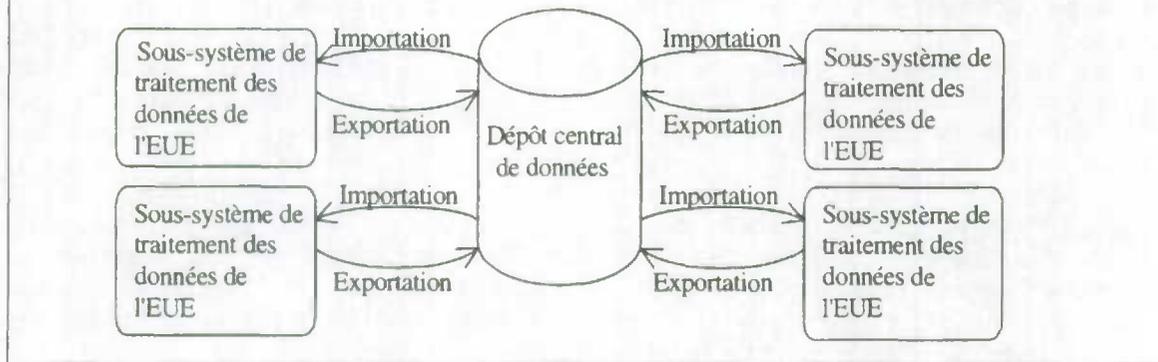
Enfin, un objectif fort important est l'amélioration permanente. Les systèmes et les méthodes que nous mettons en place doivent répondre aux exigences de la gestion en ce qui concerne les renseignements sur les processus et la transformation des données qui a lieu durant ces processus. Ce n'est que dans ces circonstances que l'on pourra prendre les décisions appropriées concernant la modification du système automatisé et l'intervention manuelle pour les futurs cycles de l'enquête.

4. Stratégie

4.1 Stratégie globale de mise en place des systèmes de l'EUE

Le sous-système de V et I est l'une des nombreuses composantes du système de traitement des données de l'EUE après la collecte. La première étape du processus de développement consistait à énoncer une conception globale des systèmes de l'EUE. La conception à laquelle sont arrivés les développeurs est celle d'un dépôt central de données (DCD), avec ses fonctions auxiliaires d'importation et d'exportation, entouré d'une série de sous-systèmes satellites, dont le système de V et I. Chaque système satellite possède son propre entrepôt de données soutenu par la technologie qui correspond le mieux aux exigences de traitement. Chaque exécution d'une composante de traitement commence par une importation de données à partir du DCD et se termine par une exportation vers ce même DCD. Les données exportées à partir de chaque sous-système sont retenues sous forme d'« instantané » dans le DCD, pour que l'on puisse retracer toutes les transformations subies au cours du cycle de traitement par les données couvrant l'année de référence complète.

Système de traitement et dépôt central de données de l'EUE



4.2 Stratégie de mise au point du système de vérification et d'imputation

Deux facteurs ont surtout influencé la conception du système de vérification et d'imputation en place à l'heure actuelle pour l'EUE, à savoir la nécessité qu'un seul système traite les données d'un grand nombre de branches d'activités fort diverses et celle de réduire au minimum les coûts de développement. La première version du système de V et I applicable à l'EUE à longue échéance devait accommoder 17 branches d'activité, 30 questionnaires et plus de 1 000 variables. Pour la deuxième année, il faudra accommoder environ 25 branches d'activité et plus de 40 questionnaires. D'autres branches d'activité seront ajoutées la troisième année, faisant passer le nombre total de questionnaires à plus de 70. Le système doit être capable de traiter des branches d'activité dont le contenu des données varie et être suffisamment souple pour permettre, d'une année à l'autre, l'ajout de nouvelles branches d'activité ou la modification du contenu des données sur les branches d'activité existantes moyennant une intervention aussi limitée que possible des spécialistes des TI.

La stratégie adoptée pour satisfaire ces critères a été de concevoir un système axé sur des métadonnées qui vise à éliminer du système de traitement toute information particulière à une branche d'activité. Essentiellement, conçu de cette façon, le processus applique une série d'algorithmes prédéfinis à un ensemble d'éléments de données abstraits. Les connaissances au sujet d'une branche d'activité particulière sont mémorisées dans une base de métadonnées que le système de traitement peut consulter au besoin.

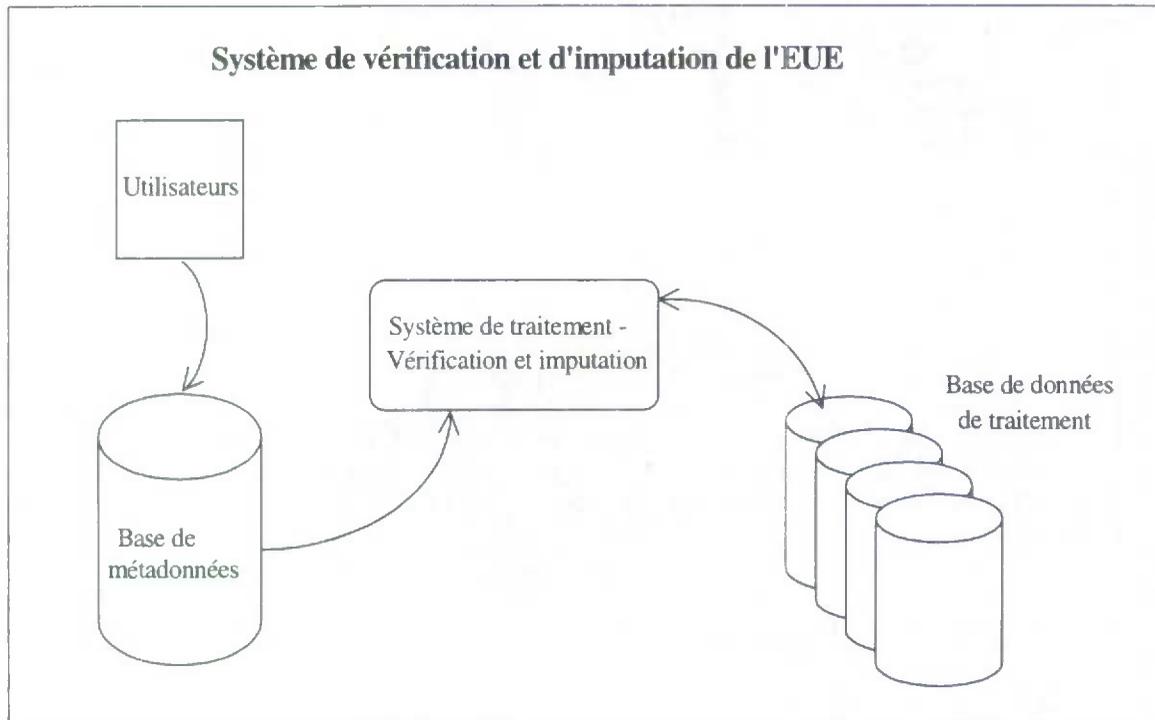
4.2.1 Base de métadonnées

La base de métadonnées satisfait trois catégories d'exigences du système. En premier lieu, elle définit le contenu des données. Elle décrit les éléments de données qui sont nécessaires pour traiter chaque branche d'activité, en ce qui concerne les données courantes, les données chronologiques, les données administratives et ainsi de suite. En deuxième lieu, elle décrit, pour chaque module ou algorithme générique, la relation entre les éléments de données auxquels il faut appliquer l'algorithme en question. Enfin, elle dicte quels modules s'appliquent et dans quel ordre il faut les appliquer.

4.2.2 Base de données de traitement

La base de données de traitement est l'entrepôt où sont stockées les données pendant leur transformation par les algorithmes du système de V et I. Pour chaque branche d'activité, le système construit une occurrence de la base de données de traitement qui contient les données

d'enquête qui doivent être traitées. Pour la plupart, on importe les variables pertinentes à partir du DCD. Après l'importation, on applique les modules génériques du système aux données d'enquête stockées dans la base de données de traitement. Puis, les résultats sont exportés vers le DCD et les tables de la base de données relatives à la branche d'activité étudiée peuvent être supprimées. Les métadonnées qui définissent le contenu pour une branche d'activité soutiennent les fonctions d'importation, d'exportation et de création de tables.



4.2.3 Système de traitement relatif à la vérification et à l'imputation

Le système de traitement relatif à la vérification et à l'imputation exécute trois grandes fonctions. Il applique les algorithmes nécessaires pour vérifier et, au besoin, transformer les données d'enquête. Il fournit les données auxiliaires au sujet des transformations. Enfin, il contrôle l'enchaînement des opérations conformément aux exigences de la branche d'activité à laquelle correspondent les données traitées.

L'élément principal est l'ensemble de modules génériques. Chacun de ces modules est entièrement indépendant du contenu des données. Chaque module est associé aux métadonnées qui soutiennent sa fonction spécifique. Par exemple, le module qui vérifie si la somme des données détaillées déclarées est égale au total déclaré a accès aux métadonnées qui permettent d'identifier les groupes de variables et leurs relations, ainsi qu'aux données qui précisent les seuils de tolérance. Le module qui repère les valeurs aberrantes qu'il faut exclure de l'ensemble des donneurs a accès aux métadonnées qui permettent d'identifier les variables à utiliser pour le calcul. Dans chaque cas, les métadonnées peuvent différer d'une branche d'activité à l'autre. Plus de trente algorithmes sont nécessaires pour satisfaire les exigences de la première version du système. On a pu utiliser plusieurs modules du SGVI pour répondre à certaines de ces exigences. Les autres algorithmes ont été créés spécifiquement pour l'EUE. Puisque le SGVI est, lui aussi, indépendant du contenu des données et axé sur des métadonnées, il s'adapte particulièrement bien au contexte de la stratégie globale.

Afin de gérer et, en dernière analyse, d'améliorer l'exécution du processus de V et I, le gestionnaire du système doit savoir quelles variables ont été transformées et par quel algorithme. A cette fin, on associe à chaque élément de donnée d'enquête qui figure dans la base de

données de traitement une variable d'état correspondante. À chaque fois qu'un élément de donnée est mis à jour, la variable d'état correspondante est modifiée de façon à indiquer l'algorithme ayant effectué la mise à jour. À la fin du processus, le gestionnaire du système de V et I peut analyser ces renseignements en vue d'évaluer le comportement de la fonction complète de V et I.

Afin de répondre aux objectifs d'exécution du système de vérification et d'imputation, l'architecture du système de traitement comprend une couche opérationnelle qui détermine quels modules sont utilisés et dans quel ordre ils sont appliqués aux données d'une branche d'activité particulière. Par conséquent, le système a été qualifié de « système à boutons-poussoirs ». Même pour les modules du SGVI, il n'est plus nécessaire de s'arrêter à chaque étape de traitement avant de passer à la suivante.

4.2.4 Le processus

Pour traiter les données sur une branche d'activité particulière au moyen du système de vérification et d'imputation :

- l'utilisateur met à jour la base de métadonnées en y ajoutant les renseignements dont le système de vérification et d'imputation a besoin pour la branche d'activité étudiée;
- d'après la base de métadonnées, le système crée une occurrence de la base de données de traitement pour la branche d'activité visée et les données sont importées à partir du DCD;
- d'après la base de métadonnées, le système applique un ensemble de modules prédéfinis à la base de données de traitement de l'industrie étudiée, puis les données sont réexportées dans le DCD.

5. Coup d'œil sur le passé

Maintenant que la première version du système proposé comme solution à long terme est en exploitation, toutes les parties concernées sont d'avis que la méthode des métadonnées est la bonne.

La base de métadonnées permet d'accommoder les éléments distinctifs des diverses branches d'activité. Un même système peut facilement traiter les données pour un grand nombre de branches d'activité, si la base de métadonnées contient des renseignements corrects sur ces branches d'activité. Pour tenir compte de la modification du contenu des données sur une branche d'activité ou de l'ajout de nouvelles branches d'activité, il suffit de mettre à jour la base de métadonnées sans devoir faire intervenir les spécialistes des TI.

Cette méthode a permis d'éliminer la duplication des systèmes de V et I. Les algorithmes ont été créés une fois pour toute, puis utilisés de nombreuses fois. La stratégie a permis de réduire les coûts initiaux de développement et les économies que l'on réalisera dans l'avenir en ce qui concerne les coûts de maintenance seront encore plus impressionnantes.

Les modules du SGVI ont été intégrés sans difficulté à un système par ailleurs complètement nouveau. Les difficultés opérationnelles posées antérieurement par le SGVI ont pu être surmontées grâce à l'ajout d'une couche opérationnelle qui prédétermine, pour chaque branche d'activité, quels modules il faut exécuter et quand. Cette stratégie assure que les échéanciers réels de traitement soient respectés, puisqu'il ne faut que peu d'interventions humaines, voire aucune, entre l'exécution des divers modules.

La décision de créer et de tenir à jour une variable d'état pour chaque variable d'enquête a été bien acceptée. Les variables d'état permettent d'exécuter des analyses à un niveau de détail que l'on ne pouvait généralement pas atteindre avec les systèmes antérieurs.

L'adoption d'une stratégie fortement modulaire offre aussi quelques avantages inattendus. Elle permet :

- de choisir quel module doit être complété en premier, puisque chaque module est entièrement indépendant des autres du point de vue de la mise en œuvre physique;
- de gérer plus facilement la mise à l'essai.

Les travaux de développement n'ont pas tous été simples et le système présente naturellement des limites.

La généralisation des modules du système de vérification et d'imputation est essentiellement axée sur le contenu. Le système peut facilement gérer les variations de contenu, mais la variation des exigences de traitement est une autre question. S'il faut procéder, pour une branche d'activité particulière, à une transformation des données que ne peut exécuter à l'heure actuelle aucun module générique, il faudra créer un nouveau module générique satisfaisant la nouvelle exigence. Toutefois, le système contient déjà un grand nombre d'algorithmes et le nombre de nouveaux modules qui seront nécessaires dans l'avenir devrait être minime.

Une autre limite importante du système tient à la base de métadonnées. Celle-ci peut être considérée comme un entrepôt contenant les connaissances sur les données recueillies pour une branche d'activité et sur la façon de traiter ces données. Afin de la tenir à jour, l'utilisateur doit bien connaître les données sur la branche d'activité étudiée, ainsi que le système de traitement. La complexité de la tenue à jour de la base de métadonnées ne doit pas être sous-estimée.

La méthode des métadonnées a été plus difficile à mettre en œuvre qu'une méthode plus classique, axée sur le domaine spécialisé, en ce qui concerne tant la programmation que la spécification des métadonnées. La courbe d'apprentissage des programmeurs a été assez abrupte et la tenue à jour des métadonnées a demandé beaucoup plus d'efforts que prévu au départ de la part des membres non techniciens de l'équipe.

L'établissement des exigences, des méthodes et des priorités du système de V et I des données de l'EUE a été une expérience douce-amère. L'année de l'essai pilote, les spécialistes du domaine représentaient sept branches d'activité qui n'avaient jamais été étudiées auparavant. Ces spécialistes étaient entièrement ouverts aux suggestions d'autres participants au projet. L'équipe est arrivée facilement à un consensus en ce qui concerne les exigences, les méthodes et les priorités. Pour la deuxième année, qui est maintenant en cours avec le nouveau système en exploitation, et pour la troisième année, pour laquelle le recueil des exigences vient d'être achevé, la tâche a été plus difficile. Dans chaque cas, au moins une branche d'activité pour laquelle existe des antécédents d'enquête a été ajoutée à l'EUE. Or, les attentes des représentants des divisions spécialisées chargés d'étudier ces branches d'activité diffèrent de celles des spécialistes des divisions chargées d'étudier les branches d'activité visées par l'essai pilote. La qualité et l'actualité des estimations produites d'après les données de l'EUE doivent égaler ou surpasser celles des résultats antérieurs. Un moyen d'assurer la constance à cet égard est de continuer à faire les choses exactement comme par le passé. Parallèlement, une nouvelle façon de faire, propre à l'EUE, est progressivement adoptée par les premiers membres de l'équipe du projet. Or, ces deux façons de voir sont souvent conflictuelles. Les tensions sont plus prononcées que durant l'essai pilote. Le consensus est plus difficile à réaliser.

6. Coup d'œil sur l'avenir

La tenue à jour de la base de métadonnées représente un tel fardeau que le développement d'une interface conviviale pour soutenir cette tâche figure très haut sur la liste des priorités. Puisque la fonction de V et I n'est pas isolée, mais fait plutôt partie intégrante d'un système plus vaste de l'EUE, le système de tenue à jour des métadonnées devra, en dernière analyse,

s'appliquer aux métadonnées de tous les sous-systèmes. Le système que l'on développera devra être intuitif et muni d'un ensemble de contrôles visant à prévenir les erreurs de nature systématique. Néanmoins, l'expérience acquise en définissant des métadonnées dans un environnement complètement manuel est inestimable. Maintenant que nous avons accompli la tâche une première fois à la dure, nous espérons faire un meilleur travail lorsque nous développerons la nouvelle interface.

Nous avons déjà précisé quels seront les nouveaux algorithmes nécessaires pour la deuxième version du système. De nouveau, certaines exigences pourront être satisfaites grâce à des modules du SGVI, tandis que d'autres nécessiteront la création de nouveaux modules.

Nous prévoyons donner encore plus de souplesse au système en ce qui concerne le choix des modules à exécuter et le moment de leur exécution. Lors de la création de la première version, nous avons, à cet égard, obtenu de meilleurs résultats pour les modules du SGVI que pour les modules nouvellement créés pour l'EUE.

Nous avons donné au système la capacité d'enregistrer des renseignements sur la transformation des données, grâce aux variables d'état. Maintenant, nous avons besoin d'outils qui nous aideront à analyser ces renseignements, afin de pouvoir perfectionner ou remanier, au besoin, les méthodes appliquées dans le système de V et I.

Série technique - Index

16 novembre, 2000

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 écrire à Statistique Canada, 13^e 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^{er} 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 Bernier 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 renonciations – 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**

**Une nouvelle stratégie de
traitement des données
pour le programme unifié
des statistiques auprès
des entreprises**

**Un exemple de vérification et
d'imputation**

Série technique

Numéro 65

**Project to Improve Provincial Economic
Statistics**

**A New Approach to
Processing for the
Unified Enterprise
Statistics Program**

**An Edit and Imputation
Example**

Technical Series

Number 65



Internet : www.statcan.ca
Intranet : <http://pasep>



Statistique
Canada

Statistics
Canada

Canada

cr905

STATISTICS CANADA LIBRARY
BIBLIOTHEQUE STATISTIQUE CANADA



1010314753

c. 2

