

## LA COLLECTE DES DONNÉES POUR L'ENQUÊTE UNIFIÉE SUR LES ENTREPRISES

Sylvain Perron, Mike Sirois, Jean Drouin et Sylvain Poirier<sup>1</sup>

### RÉSUMÉ

Dans le cadre du Projet d'amélioration des statistiques économiques provinciales (PASEP), une nouvelle enquête unifiée sur les entreprises (UES) est mise sur pied. À long terme, cette enquête couvrira tous les secteurs industriels de l'économie canadienne. Par contre, pour la première année, l'enquête sera menée auprès de répondants de sept secteurs industriels qui n'ont jamais été couverts par Statistique Canada auparavant ou qui n'ont pas été couverts récemment. Une stratégie de collecte a donc été développée pour recueillir les données de cette enquête unifiée. Dans cet article on verra tout d'abord le contexte dans lequel cette nouvelle stratégie a été élaborée. On poursuivra avec une description de la stratégie de collecte et des systèmes qu'elle implique. On verra ensuite la stratégie de vérification et de suivi adoptée. On terminera avec un bref aperçu du travail qui sera entrepris dans les années futures lorsque l'enquête unifiée inclura d'autres secteurs économiques.

**MOTS-CLÉS** : Interview téléphonique assistée par ordinateur; règles de vérification; prioriser; stratégie de test.

### ABSTRACT

Within the framework of the Project to Improve Provincial Economic Statistics (PIPES) a new Unified Enterprise Survey (UES) is being developed. This survey will eventually cover all the industrial sectors of the Canadian economy. However, for the first year, the UES will be sent to enterprises from 7 industrial sectors never surveyed before by Statistics Canada, or not covered by a survey recently. A data collection strategy has been developed for the UES. In this paper, we will first describe the context in which the strategy was elaborated. We will then describe the strategy and the systems it requires. This will be followed by a description of the follow-up and editing strategy that will be used in the data collection process. We will finish by giving an overview of the work that will be done in the coming years, as the UES comes to encompass all industrial sectors.

**KEY WORDS** : Computer assisted telephone interview; Edit rules; Prioritize; Testing strategy.

### 1. INTRODUCTION

Dans le cadre de l'enquête unifiée sur les entreprises (EUE), la stratégie globale de collecte des données adoptée consiste à faire un envoi postal des questionnaires, pour ensuite faire un suivi téléphonique. Ce suivi se fera à partir d'un système d'interview téléphonique assistée par ordinateur (ITAO). L'objectif du système de collecte est de recueillir les données auprès des répondants, d'effectuer les corrections et les suivis nécessaires afin que les données soient les plus correctes possibles, et enfin de transmettre les données et l'information pertinente au système de traitement qui suit le processus de collecte.

Si on revient à l'envoi des questionnaires, il faut décider comment et à qui ceux-ci seront envoyés. En effet, comme l'EUE s'adresse aux entreprises et aux établissements, on doit définir l'unité de collecte c'est-à-dire l'unité qui recevra les questionnaires. L'unité de collecte ne correspond pas nécessairement à une entreprise ou à un établissement, il peut s'agir par exemple d'un agrégat d'établissements. Il faut donc s'adapter aux répondants, et permettre ce que nous appelons des rapports unifiés. Ces rapports unifiés permettent à un répondant de répondre pour toutes ses entreprises et établissements. Ce répondant recevra ainsi tous les questionnaires pertinents.

De plus, nous avons aussi défini des unités de collecte spéciales pour les entreprises qui sont problématiques

<sup>1</sup> Sylvain Perron, Mike Sirois, Jean Drouin, Sylvain Poirier, Division des Méthodes d'enquêtes-entreprises, Statistique Canada, Ottawa, Ontario, K1A 0T6

par leur complexité (elles ont de très grands revenus, un très grand nombre d'établissements, etc.). Il y a aussi les entreprises qui ont des arrangements spéciaux avec Statistique Canada. Ces unités seront contactées par des gestionnaires des répondants clés (GRC), des personnes ressources spécifiquement assignées à ces entreprises. Nous ne verrons pas en plus grands détails la définition des unités de collecte.

Dans cet article, nous allons maintenant voir les types de questionnaires qui existent au niveau de l'EUE. Nous passerons ensuite aux règles de vérification développées pour les applications de collecte. Nous verrons par la suite la stratégie de suivi adoptée pour l'EUE. Puis, nous passerons à la stratégie de tests des applications. Nous terminerons enfin avec une brève conclusion et un aperçu du travail à venir.

## **2. QUESTIONNAIRES**

Il y a deux types de questionnaires dans le cadre de l'EUE, les questionnaires destinés aux entreprises, et ceux destinés aux établissements. Pour chacun des types de questionnaires, une stratégie de collecte différente sera utilisée, toujours en respectant le principe de l'envoi postal et du suivi téléphonique.

### **2.1 Questionnaire entreprise**

Il y a quatre questionnaires différents pour les entreprises. Ce sont les questionnaires destinés aux entreprises du secteur financier et ceux destinés aux entreprises du secteur non financier. Pour chacun, nous avons un questionnaire court ou un questionnaire long. Afin de recueillir les données des questionnaires entreprises nous allons utiliser un système de collecte déjà existant à Statistique Canada.

Quant à la différence entre les questionnaires court et long, une entreprise recevra un questionnaire court si elle fait partie de l'Enquête trimestrielle de la Division de l'organisation et des finances de l'industrie (DOFI). Dans ce cas, nous avons déjà recueilli une bonne partie des données requises par l'EUE. Par conséquent, le questionnaire qui sera envoyé à cette entreprise sera un questionnaire court. Dans le cas contraire, nous devons recueillir l'information complète auprès du répondant, d'où le questionnaire long. Pour les fins de cet article, nous n'approfondirons pas plus en détails le questionnaire entreprise.

### **2.2 Questionnaire établissement**

En ce qui concerne les questionnaires établissement, on peut séparer les questionnaires selon qu'ils s'adressent aux entreprises complexes ou aux

entreprises simples. Au niveau des entreprises complexes, nous avons sept questionnaires qui couvrent les sept secteurs industriels de l'enquête pilote (l'enquête de la première année). Il y a aussi un questionnaire spécial qui sera envoyé aux sièges sociaux des entreprises. Donc, chacun des établissements appartenant à une entreprise complexe recevra un questionnaire si l'établissement appartient à un des secteurs industriels de l'enquête pilote. Les autres établissements seront ignorés. De plus, les établissements qui ont de petits revenus ne recevront aucun questionnaires, seules les données fiscales seront utilisées.

Pour les entreprises simples, nous avons des questionnaires longs et des questionnaires courts. Les établissements des entreprises simples qui appartiennent aux sept secteurs industriels de l'enquête pilote, recevront un questionnaire long ou un questionnaire court selon leurs revenus. On utilisera les données fiscales pour compléter les données d'enquêtes des établissements qui recevront des questionnaires courts. De même que pour les établissements des entreprises complexes, les établissements des entreprises simples qui ont de petits revenus ne recevront aucun questionnaires, seules les données fiscales seront utilisées.

Pour l'enquête pilote, au niveau des questionnaires établissements, les questionnaires des entreprises complexes et les questionnaires longs des entreprises simples seront exactement les mêmes à cause de plusieurs contraintes, dont des contraintes de temps. Cette stratégie est appelée à changer dans le futur. De plus, il n'y aura que deux secteurs (industrie de la construction et industrie des services de restauration et des débits de boisson) qui auront un questionnaire court.

## **3. SYSTÈME DE COLLECTE ET RÈGLES DE VÉRIFICATION**

### **3.1 Système de collecte**

Comme nous l'avons mentionné à la section 2.1, les questionnaires entreprises seront traités avec un système de collecte déjà existant, il s'agit maintenant de concevoir le système de collecte pour les questionnaires établissement seulement. Pour ce faire nous utiliserons le système CASES, un système conçu pour développer des applications d'entrevue téléphonique assistée par ordinateur.

Pour nos besoins, parce que les questionnaires sont très différents entre eux, nous devons bâtir une

application différente pour chaque questionnaire. Chacune des ces applications contiendra les questions et les règles de vérification appropriées. Ces applications seront utilisées pour faire à la fois les suivis téléphoniques et la saisie des questionnaires retournés par la poste. Il y aura en tout neuf applications, sept pour les questionnaires longs, et deux pour les questionnaires courts.

Au niveau du contrôle de la qualité, l'utilisation d'un moniteur témoin nous permettra de suivre des entrevues et de voir en même temps l'entrée des données par l'interviewer. Cette stratégie initiale sera améliorée au cours des années.

### 3.2 Règles de vérification

Avec les différents questionnaires, les règles de vérification sont un élément essentiel de la programmation des applications CASES. Puisque les sept secteurs industriels choisis pour l'enquête pilote sont des secteurs nouvellement enquêtés, nous ne possédons pas de données nous permettant de faire des comparaisons historiques. Il n'y aura donc pas de règles historiques dans les applications. Par conséquent, il y aura quatre types de règles pour la collecte, soient :

La règle d'étendue : (borne1 < X < borne2)

La règle de validation : (A + B + C = total) avec une marge d'erreur

La règle de cohérence : (si A > 0 alors B > 0)

La règle relationnelle : (borne1 < X/Y < borne2)

Comme son nom l'indique, la règle d'étendue permet de vérifier que la valeur que prend une variable se trouve entre deux paramètres bien définis. Par exemple, les revenus donnés par un répondant doivent être supérieurs à zéro et inférieurs à 999 999 999 \$. Dans le cas contraire, la règle échoue.

La règle de validation permet quant à elle de comparer les valeurs correspondant aux composantes d'un total avec la valeur correspondant à ce total. On s'assure que la différence entre la somme des composantes et la valeur du total est nulle (dans ce cas la somme correspond exactement au total) ou comprise à l'intérieur d'une marge d'erreur. La marge d'erreur permet de tenir compte des erreurs d'arrondi ou des erreurs mineures. Pour la première année, la marge d'erreur a été fixée à 2%.

C'est par la règle de cohérence qu'on s'assure de la présence d'une réponse à une variable lorsqu'une autre variable est présente dans le questionnaire. Par exemple, si un répondant indique qu'il a des revenus dans la catégorie « autres revenus », on s'assurera que la variable correspondant au champ « s.v.p. spécifiez » n'est pas vide. Le répondant doit nous indiquer la nature de ses autres revenus.

Enfin la règle relationnelle permet de vérifier les liens entre les valeurs de deux variables. Contrairement à la règle d'étendue, où c'est la valeur d'une variable qui doit se trouver entre deux bornes, ici on compare le ratio entre deux variables. C'est ce ratio, par exemple le ratio des revenus par rapport aux traitements et salaires, qui doit se retrouver entre deux bornes spécifiées dans l'application.

Afin d'aider lors de la création des règles de vérification, et de s'assurer que toute l'information requise pour pouvoir les programmer est présente, un document a été écrit (Drouin et Perron 1998). Ce document indique entre autres comment les règles doivent être spécifiées et quelle information est nécessaire et de quelle façon on doit la donner. On indique aussi quelles sont les questions à poser aux répondants lorsqu'une règle de vérification échoue. Ce document deviendra donc la pierre angulaire pour la spécification des règles de vérification. Il sera amélioré au fil des années.

### 3.3 Information transmise au système de traitement

Le système de traitement des données suit l'étape de collecte. Il est conçu afin de recevoir non seulement les données recueillies auprès des répondants lors de l'étape de collecte, mais aussi de l'information supplémentaire générée par les applications CASES. Cette information est transmise à l'aide de variables indicatrices et donne essentiellement ce qui s'est passé lors de la collecte des données.

Alors qu'habituellement nous avons qu'un statut global pour l'état du questionnaire (complet, refus, etc.), nous aurons un statut pour chaque variable du questionnaire jugée importante. Nous aurons donc, pour chacune de ces variables, de l'information pertinente sur les règles de vérification appliquées à la variable. Nous aurons le statut de la règle, valeur échouée, valeur confirmée par le répondant, règle toujours en erreur, refus de répondre, ne sais pas, etc. Cette information sera utilisée lors de la vérification et de l'imputation. On tiendra compte du fait qu'une variable est confirmée par le répondant, et on ne la

changera pas par l'imputation même si la variable échoue des règles de vérification.

#### 4. STRATÉGIE DE SUIVI

La stratégie de suivi qui a été développée pour les applications de collecte s'applique autant pour la collecte d'un questionnaire papier retourné par la poste, que pour le suivi d'un non répondant. En effet, le système de collecte est utilisé pour faire la saisie des données des questionnaires retournés par la poste. La stratégie développée est assez élémentaire pour le projet pilote. Elle sera raffinée au cours des années.

Les objectifs du suivi sont en premier lieu de suivre les questionnaires les plus importants. On veut donc suivre les répondants ayant le plus d'impact sur les estimations, tout en respectant les autres objectifs. En second lieu, on doit permettre de bâtir les bassins de donneurs requis par le système de traitement (pour l'imputation). Comme le système d'imputation requiert un nombre minimum de donneurs par groupes d'imputation (secteurs industriels par province), il faut s'assurer que les suivis nécessaires sont faits par secteurs industriels et province.

Enfin, comme troisième objectif, la stratégie de suivi doit être équitable pour chacun des secteurs industriels puisque les taux d'imputation seront considérés lors du calcul des coefficients de variation pour les estimations. Comme la variance due à l'imputation sera un facteur dans le calcul de la variance totale, et que cette variance est influencée par la non-réponse, il faut s'assurer d'être équitable avec chaque secteur.

La stratégie consiste donc à prioriser les questionnaires dans un ordre respectant nos objectifs. Cet ordre de priorité sera défini selon l'importance des questionnaires et selon les règles de vérification échouées. Pour ce faire, nous allons développer une matrice qui comprendra d'une part les facteurs reliés à l'unité, et d'autre part les facteurs reliés aux règles de vérification. Au niveau de l'unité on tiendra compte de la taille de celle-ci, de sa strate (classification industrielle) et enfin de son type complexe ou simple. En ce qui a trait aux règles de vérifications, on se préoccupera de savoir s'il y a absence d'une des 3 variables d'appariement (revenu total, dépenses totales, nombre d'employés), s'il y a absence d'une des 3 variables importantes (période rapportée, salaires, classifications industrielles), si les règles échouées comprennent des ratios aberrants ou enfin s'il y a des règles échouées.

Cette matrice permettra d'établir l'ordre dans lequel les suivis seront faits. Lors d'un suivi, on tentera de corriger toutes les erreurs d'un questionnaire. Puisqu'on est en contact avec le répondant, on corrige tout le questionnaire.

De plus, comme il a été mentionné auparavant, l'information concernant les résultats des règles de vérification et des suivis sera transmise au système de traitement qui suit le processus de collecte. Pour chaque variable importante d'un questionnaire, on aura si la variable a échoué une règle de vérification, est-ce que c'est une valeur corrigée, une valeur confirmée, y a-t-il eu un refus de répondre de la part du répondant, ou est-ce qu'il ne sait pas. Toute information jugée pertinente va être transmise afin que le système de vérification et d'imputation respecte le plus possible les données fournies et confirmées par le répondant.

#### 5. STRATÉGIE DE TEST DU SYSTÈME

Une fois les applications de collecte des données programmées, il faut s'assurer qu'elles fonctionnent correctement. Nous devons ici tester deux aspects en particulier : les applications elles-mêmes et aussi les fichiers de sortie qui seront transmis. Une stratégie de test a été mise sur pieds. Cette stratégie vise à répondre aux deux points, et tente de s'assurer que les tests des systèmes seront complets.

Au niveau de la vérification d'une application, il faut s'assurer que le déroulement de celle-ci est correct et suit le questionnaire. Il faut aussi vérifier que les règles de vérification ont été bien programmées, et que les branchements aux questions suivantes ou encore au suivi se font bien. Il faut aussi vérifier que les questions utilisées pour le suivi des règles échouées sont correctes, arrivent dans le bon ordre, et ramènent l'interviewer au bon endroit dans le questionnaire.

Lorsqu'il s'agit de tester les fichiers de sortie, il faut s'assurer que les variables appropriées sont écrites dans le fichier. En plus, il faut s'assurer du positionnement, de la valeur et du format des variables. Enfin, on doit s'assurer que les variables indicatrices (qui contiennent l'information pertinente aux variables importantes) sont-elles aussi présentes, bien positionnées et avec le format adéquat.

Pour aider avec la vérification, deux documents correspondant à la stratégie ont été écrits (voir Poirier 1998a et 1998b). Le premier document, intitulé «Testing strategy of the capture edit specifications version 1.0», donne une stratégie globale de test pour les applications de collecte. Le second, «Testing the

output file from the data collection version 1.0», correspond à la vérification des fichiers de sortie.

## 6. CONCLUSION ET TRAVAIL FUTUR

En guise de conclusion, il faut mentionner que malgré de sévères contraintes, nous avons réussi à mettre sur pieds toutes les applications de collecte nécessaires pour l'année pilote. Il y a eu beaucoup de travail accompli. La mise à la poste des questionnaires, une opération très complexe a été très bien réussie. Nous avons acquis beaucoup d'expertise, ce qui sera très important pour les années qui viennent.

Au cours de ces années, l'intégration de nouveaux secteurs industriels sera au cœur de tout le travail qui sera à accomplir. Parmi les autres tâches à faire, on peut mentionner l'inclusion de règles de vérification historiques à l'intérieur des applications de collecte des données. On envisage aussi d'utiliser un nouveau système de saisie pour développer les applications de collecte. Ce système devrait permettre avec plus de facilité la saisie rapide des questionnaires retournés par la poste.

De plus, cette année nous avons développé un système de base qui permet aux répondants de fournir ses données de façon électronique. Pour le moment ce système est élémentaire et doit faire face aux problèmes de confidentialité. Par contre, il sera amélioré, et le nouveau système de saisie devrait nous

permettre plus facilement l'inclusion des données des répondants électroniques.

Enfin, différentes avenues doivent être développées et offertes aux répondants. Parmi celles-ci, on retrouve la mise à la poste échelonnée afin de contacter les répondants au moment le plus pertinent pour lui (par exemple peu après la fin de son exercice financier). Il y a aussi la création de questionnaires personnalisées, qui ne contient que les questions pertinentes pour le répondant (par exemple une liste de commodités adaptée au répondant). Ces développements visent à faciliter la tâche du répondant, et donc améliorer les résultats de l'enquête.

## RÉFÉRENCES

Jean Drouin et Sylvain Perron (1998), Creation of the collection specifications for the CASES applications in the context of the Unified Enterprise Statistics Program (UESP) version 1.1, Document de travail, *Division des méthodes d'enquêtes entreprises, Statistique Canada.*

Sylvain Poirier (1998a), Testing strategy of the capture edit specifications version 1.0, Document de travail, *Division des méthodes d'enquêtes entreprises, Statistique Canada.*

Sylvain Poirier (1998b), Testing the output file from the data collection version 1.0, *Division des méthodes d'enquêtes entreprises, Statistique Canada*